



مركز البحوث والدراسات

ثورة البيانات

البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، والنتائج المترتبة عنها

تأليف
بروفيسور: روب كيتشن

ترجمة: د. محمد بن أحمد غروي
راجع الترجمة: ماجد بن عبدالعزيز البريثن





مركز البحوث والدراسات

ثورة البيانات

البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، والنتائج المترتبة عنها

تأليف

بروفيسور: روب كيتشن

ترجمة: د. محمد بن أحمد غروي

راجع الترجمة: ماجد بن عبدالعزيز البريثن

١٤٣٩هـ - ٢٠١٨م



بطاقة الفهرسة

ح معهد الإدارة العامة، ١٤٣٩هـ.

فهرسة مكتبة الملك فهد الوطنية أثناء النشر.

ثورة البيانات المفتوحة والبنى التحتية للبيانات، والنتائج المترتبة عنها / روب كيتشن، محمد بن أحمد غروي، ماجد بن عبدالعزيز البريش - الرياض، ١٤٣٩هـ

٣٦٤ص؛ ١٧سم × ٢٤سم.

ردمك: ٥-٢٦٥-١٤-٩٩٦٠

١- الإنترنت أ- غروي محمد بن أحمد (مترجم) ب- البريش، ماجد بن عبدالعزيز (مراجع) ج- العنوان.

ديوي ٠٠٤,٦٧   ١٤٣٩/١٢٧٣

رقم الإيداع: ١٤٣٩/١٢٧٣

ردمك: ٥-٢٦٥-١٤-٩٩٦٠



هذه ترجمة لكتاب:

THE
DATA
REVOLUTION
BIG DATA, OPEN DATA INFRASTRUCTURES & THEIR CONSEQUENCES

Rob Kitchin



حول المؤلف
البروفيسور روب كيتشن

هو باحث رئيسي في المجلس الأوروبي للبحوث المتقدمة (European Research Council) في جامعة إيرلندا الوطنية في ماينوث. وقد قام بتأليف وتحرير كتب عديدة أخرى ناهزت ثلاثة وعشريـن كتاباً، وكان قد حاز على الميداليـة الذهبية للعلوم الاجتماعية من الأكاديمية الملكية الأيرلنديـة للعـام ٢٠١٣. وهو الباحث الرئيـسي للحاوية الرقمية الأيرلنديـة والمرصد البحثي لعموم الجزيرة الأيرلندية.



قائمة المحتويات



قائمة الجداول

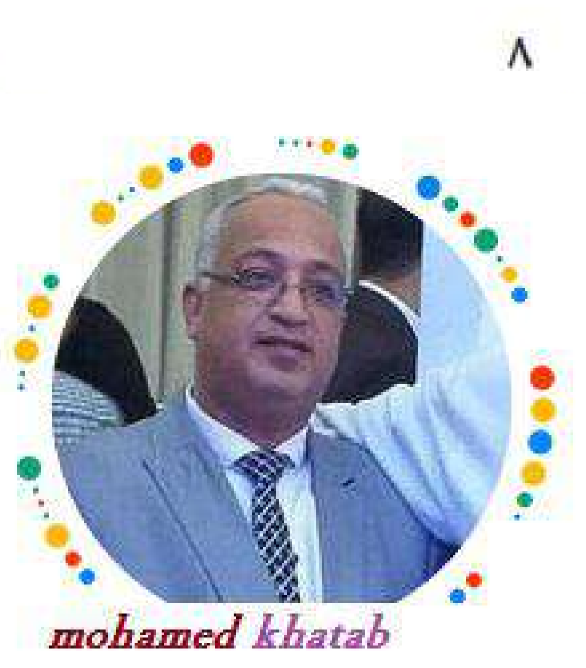


قائمة الأشكال الرسومية


mohamed khatab

mohamed khatab

mohamed khatab


mohamed khatab

mohamed khatab

mohamed khatab

mohamed khatab

mohamed khatab



شكر وتقدير:

بدأت نشـأة هذا الكتاب في أوائل الشـهر السـابع (يوليو/ آيار) من العام ٢٠١٢ خلال مناقشـة بمقهـى في أدنبرة مـع روبرت روجـك (Robert Rojek) من شركة سـايج للنشر (SAGE). حيـث كنت قد اقترحت عليه البحث عن شـخص لكتابـة مصنّف عن البيانات الكبـيرة، والبيانات المفتوحة، والبنـى التحتية للبيانات، مع تقديم الأفكار وإسـداء النصح عن ذلك الشـخص الذي قد يكون مناسباً لصياغة النص في مثل هذا الموضوع. وكان يرى في حينه أنّي ذلك الشـخص المناسب لتلك المهمة. فقررت في وقت لاحق، بعد مضي عدة أشهر، الموافقة بين خطط كتاباتي والجمع بينها، بحيث بدأت في صياغة ما بدا في متناول اليد خلال فترة زمنية قصيرة جداً ويتناول التحليل النقدي لمشهد البيانات المتغير باطراد. تطور الكتاب مع مرور الوقت إلى مخطوط كامل يسـعى إلى الدراسة المستفيضة للاتجاهات والمناقشات الناشئة بتأنّ وإنصاف. وبالتزامن مع ذلك، بقي روبرت متابعاً متحمساً ومصدراً للعديد من المواد المثيرة للاهتمام، وكانت مساعدته موضع تقدير كبير جداً. وفي سايج أيضاً، ساعد زميله كيري ديكنز (Keri Dickens) في الدفع بالكتاب إلى عملية الإنتاج والنشر، التي تمّ توجيهها على نحو مثير للإعجاب من قبل كاثرين هاو (Katherine Haw).

قـام مارتـن دودج (Martin Dodge) وتريـسي ب. لاوريولـت (Tracey P. Lauriault) بقـراءة مفصلة ونقد هـادف للمخطوط بالكامل وعلى نحو متناه في السلاسـة واللطف. وقام مـارك بويل (Mark Boyle) بقراءة كامل المسـودة الثانية، كما قدّم غافـن مكاردل (Gavin McArdle) وإيفلـين روبيرت (Evelyn Ruppert) نقـداً مفيداً لبعض فصول الكتاب، وإضافة لذلك شارك عدد من الزملاء والأصدقاء في مداخلات نافعة ومناقشات مفيدة، حيث دلّني العديد منهـم إلى مواد ذات صلة، منهم مـارك غراهام (Mark Graham)، تايلور شـيلتون (Taylor Shelton)، مات زوك (Matt Zook)، مات ويلسـون (Matt Wilson)، ليف مانوفيتش (Lev Manovich)، سـيان اوكالاهان (Cian O'Callaghan)، سـونغ - يـوه بيرنج (Sung-Yueh Perng)، أيلـين أوه كارول (Aileen O'Carroll)، جـين جراي (Jane Gray)، سـاندرا كولينز (Sandra Collins)، جون كيتنغ (John Keating)، شـارون ويب (Sharon Webb)، جوسـتين



جليسون (Justin Gleeson)، أويف داولنغ (Aoife Dowling)، يوغان مكارثي (Eoghan McCarthy)، مارتن تشارلتون (Martin Charlton)، تيم مكارثي (Tim McCarthy)، جان ريجبي (Jan Rigby)، روب برادشو (Rob Bradshaw)، آلان مور (Alan Moore)، دراش ماك دونشا (Darach Mac Donncha)، وجيم وايت (Jim White). كما تلقيت ردوداً مفيدة أيضاً خلال العروض التقديمية في جامعة دورهام، وجامعة كلارك، وجامعة هارفارد. واهتمت رونا برادشوو (Rhona Bradshaw) وأورلا دان (Orla Dunne) بإدارة أعمال المكتب عندما كنت أحاول الانكباب على إجراء البحوث وكتابة الفصول. كما تفضلت جوستين جلسيون (Justin Gleeson) بإنتاج بعض الرسوم البيانية. فأنا مدين لكم جميعاً بالعرفان. كما أود أيضاً شكر الكثير من الناس المشاركين على موقع التواصل الاجتماعي - التويتر وتغريداتهم التي أشارت إلى مواد مثيرة للاهتمام، وانخراطهم في مناقشات علمية دقيقة ذات صلة. وأخيراً، الشكر موصول إلى كورا (Cora) التي قامت بتثبيتي وقدمت دعماً رائعاً لي.

وتمَّ دعم جزء من الأبحاث التي أجريت في كتابة هذا الكتاب من خلال جائزة المجلس الأوروبي للبحوث المتقدمة، "المدينة المبرمجة" (ERC-2012-AdG-323636; www.nuim.ie/progcity) والتمويل المقدم من هيئة التعليم العالي في برنامج البحوث في المستوى الثالث للدورة الخامسة، والذي يهدف إلى إنشاء حاوية رقمية لأيرلندا.

ويمكن الاطلاع على نسخة من مراجع الكتاب من خلال الرابط التالي: http://thedatarevolutionbook.wordpress.com. كما يجري بانتظام إضافة مصادر للمعلومات وتجارب حول ثورة البيانات على الرابط http://www.scoop.it/t/the-programmable-city. وبالطبع، فإن التعليقات والردود مرحب بها من خلال البريد الإلكتروني Rob.Kitchin@nuim.ie أو التغريد على حساب موقع التواصل الاجتماعي - تويتر @robkitchin.

لقد سبق نشر بعض المواد الموجودة في هذا الكتاب على شكل أوراق بحثية أو على النشرات اليومية، على الرغم من أنه قد تم تحديثها، وإعادة صياغتها، والتوسع فيها. ومنها:



- Dodge, M. and Kitchin, R. (2005) 'Codes of life: identification codes and the machine-readable
- world', Environment and Planning D: Society and Space, 23(6): 851–81.
- Kitchin, R. (2013) 'Big data and human geography: opportunities, challenges and risks', Dialogues in
- Human Geography, 3(3): 262–7.
- Kitchin, R. (2014) 'The real-time city? Big data and smart urbanism', GeoJournal 79(1): 1–14.
- Kitchin, R. (2014) 'Big data, new epistemologies and paradigm shifts', Big Data and Society, 1(1)
- April–June, 1–12.
- Kitchin, R. and Lauriault, T. (2014) Small Data, Data Infrastructures and Big Data. The
- Programmable City Working Paper 1. Available at SSRN: http://ssrn.com/abstract=2376148.
- Kitchin, R. and Lauriault, T. (in press) 'Small data in an era of big data,' Geo Journal.

- Figure 1.1 is adapted from InformationisBeautiful.net with the permission of David McCandless.
- Figure 1.2 is reproduced with the permission of The Statistical and Social Inquiry Society of Ireland.
- Table 2.4 is included with the permission of Neil Beagrie, Brian Lavoie and Matthew Woollard and
- under a creative commons licence for Fry et al., http://repository.jisc.ac.uk/279/.
- Table 3.1 is reproduced from http://opendefinition.org/od/ under a creative commons licence.



- Table 3.3 is included with the permission of Michael Hausenblas, http://5stardata.info/.
- Table 4.1 is reproduced with the permission of The Economist. The Economist Newspaper Limited,
- London, issued March 11, 2014.
- Figure 6.1 is reproduced with the permission of Monica Stephens.
- Table 6.1 is reproduced with the permission of Taylor and Francis.
- Figure 6.2 is reproduced with the permission of Flightradar24.com.
- Figure 6.3 is reproduced with the permission of Andrew Hudson-Smith.
- Figures 6.4 and 6.5 are reproduced with the permission of Professor Mikael Jern, National Center for
- Visual Analytics, Linköping University, http://ncva.itn.liu.se.
- Table 7.1 Forms of big data corporate intelligence is included with the permission of McKinsey &
- Company.
- Table 7.2 and Figure 7.1 are reproduced courtesy of International Business Machines Corporation, ©
- International Business Machines Corporation.
- Figure 7.2 is reproduced from http://ipprio.rio.rj.gov.br/centro-de-operacoes-rio-usa-mapas-feitospelo-
- ipp/ under a creative commons license.
- Tables 10.2 and 10.3 are included with the permission of John Wiley & Sons.
- Table 10.4 is included with the permission of Ann Cavoukian, Ph.D., Information and Privacy Commissioner, Ontario, Canada.



ملاحظة:

يُعــبَرَ عــن مصطلح البيانــات (Data) في هــذا الكتاب في صيغة الجمع، مع اســتخدام مصطلح وحدة بيان (Datum) للدلالة على المثل المفرد (Singular Instance). بحسب ما هو موضح في قاموس أوكسفورد للغة الإنجليزية (Oxford English Dictionary- OED):

في اللغة اللاتينية، البيانات هي صيغة جمع لكلمة بيان، وتاريخياً وفي المجالات العلمية المتخصصــة، على الخصوص، يتمَ التعامل أيضاً مــع مصطلح «البيانات» للدلالة على الجمع في اللغــة الإنجليزية، حيث يأتي في ســياق فعل الجمع (were)، كــما في «تمَ جمع البيانات وتصنيفها».

ومع ذلك، يســتخدم هذا المصطلح على نحو متزايد في صيغة المفرد في وســائل الإعلام العامة والمحادثة اليومية، كما جرى تفصيل ذلك في قاموس أوكسفورد للغة الإنجليزية:

ومــع ذلــك، في العموم لا يجــري التعامل مع مصطلــح «البيانات» بصيغــة الجمع في الاســتخدام غير العلمي الحديث، ولكن بدلاً من ذلك، يجرى التعامل معه كاســم إجمالي - للبنــد غير المعدود – على غــرار كلمة «معلومات»، التي تأخذ الفعل بصيغة المفرد. وأصبح من المقبول في اللغة الإنجليزية القياســية على مدى عدد من الســنوات وعلى نطاق واسع الجمل من قبيل: «تم جمع البيانات» في سياق فعل المفرد «was».

لذا، يتبع هذا الكتاب الاصطلاح العلمي المعمول به. ومع ذلك، تم إبقاء النص على حاله حيثما اتفق ورود استخدامه بصيغة المفرد عند الاقتباس. ودون الجدل في أي الصيغتين هو الأصح، فإن انحياز النحاة هو لصيغة فعل الجمع، ولكن الاســتخدام العام الشــائع هو أكثر انفتاحاً ومرونة.


المكتبة الاقتصادية
قناة المكتبة على التيليجرام
ECONLIBRARY


المكتبة الاقتصادية



مقدمة:

هناك تاريخ طويل للحكومات، والأعمال، والعلوم، والمواطنين فيما يتعلق بإنتاج واستخدام البيانــات من أجل رصد العالم، وتنظيمه، والانتفاع منه، وجعله أكثر منطقاً وذا مغزى. وقد كانت البيانات في العادة مكلفة وتستغرق وقتاً طويلاً في الإنتاج والتحليل والتفسير وتوفر في العموم صورة ثابتة، غالباً مشوشة، وجزئية عن الظاهرة قيد الدراسة. ونظراً للندرة النسبية، كانت البيانات ذات الجودة العالية سلعة ذات قيمة، وعلى هذا الأساس كان التكتم الشديد والحرص عليها وتداولها وتبادلها بأثمان باهظة. ولكن هذا الوضع بدأ يتغير بشــكل جذري تمامــاً. فالبيانات لم تفقد شــيئاً من قيمتهــا، ولكن جرى في نواح أخــرى تحول في طبيعتها وطــرق إنتاجها من خلال مجموعة ما وصفهــا Christensen (1997) بالابتكارات المربكة، والتي شــكلت تحدياً للوضع الراهن في كيفية إنتاج البيانــات وإداراتها وتحليلها وتخزينها والاستفادة منها. فبدلاً من أن تكون البيانات نادرة ومحدودة في قدرة الوصول إليها، غدت على نحو متزايد طرق إنتاج البيانات كالطوفان؛ ســيلاً عميقاً وواســعاً من البيانات المتنوعة في طبيعتها، والمرتبطة منطقياً، والمتوافرة عند الحاجة إليها في الوقت المناســب، والمنخفضة نسبياً في التكلفة، وخارج نطاق الأعمال في طبيعتها، مع تنامي فتحها وازدياد قابلية وسهولة الوصــول إليهــا. فثورة البيانات جاريــة في الوقت الراهن، وهي تعمــل بالفعل على إعادة تشكيل كيفية إنتاج المعرفة، وإجراء الأعمال، وحوكمة التشريعات والقوانين.

لقد تأسســت هذه الثورة في الموجة الأخيرة من تقنيات المعلومات والاتصالات (ICTs)، وذلــك من خلال عدد كبير مــن الأجهزة الرقمية التــي نصادفها اليوم في المنــازل، وأماكن العمل، والأماكن العامة؛ كأجهزة الاتصال المحمولة، والحوسبة الموزعة، والحوسبة السحابية؛ ووســائل التواصل الاجتماعي؛ وإنترنت الأشــياء (أجهزة ووسائل استشعار مترابطة ومتصلة من خلال الشــبكة العالمية -الإنترنت). فهذه الوسائط والمنصات التقنية الجديدة أضحت تقود وتوجه أكثر من أي وقت مضى المزيد من جوانب الحياة اليومية – العمل، والاستهلاك، والســفر، والاتصال، والترفيه – وأضحت العوالم التي نعيش فيها تُفهم بهيئة بيانات وتُســير من خلال التقنيات المعتمدة على البيانات. وعلاوة على ذلك، فإن هذه الوســائط والمنصات التقنيــة الجديــدة تعمل مادياً ومنطقياً على إعادة تشــكيل طرق إنتاج، وتداول، وتفســير



البيانات، وإنتاج ما اصطلح على تسميته «البيانات الكبيرة Big data»، حيث كميات هائلة مـن البيانات الرقمية المتغيرة والمتنوعة، والتي من السـهولة بمـكان الجمع بينها، وتبادلها، وتوزيعها من خلال شبكات تقنية المعلومات والاتصالات، وتحليلها بواسطة جيل جديد من أسـاليب تحليل البيانات (Data Analytics) المعدَّة لتتلاءم مع وفرة البيانات على العكس مـن ندرة البيانات المعتادة. ويتضح حجم طوفان البيانات الناشـئة من خلال الادعاء بأنه "مـن فجر الحضارة وحتى العام ٢٠٠٣، أنتجت البشرية خمسـة إكسـا بايت (Exabyte) من المعلومات فقط؛ في حين نحن بصدد إنشـاء القدر نفسه من البيانات كل يومين" (Hal Varian, chief economist with Google, مقتبس من Smolan and Erwitt 2012).

ولا تُعـد البيانات الكبيرة المكـون الوحيد لثورة البيانات، فبالأحـرى كان هناك العديد من المبـادرات المتصلة بعضهـا ببعض من قبيل التحـول الرقمـي (Digitization)، ربط وتوسـيع مجموعات البيانات المنتجة تقليدياً (البيانات الصغيرة) مع بعضها عبر بنى تحتية شـبكية؛ وحركة البيانات المفتوحة التي تسـعى لجعل أكبر قدر ممكن من البيانات متوفرة بشـكل مفتوح للجميع للاستخدام؛ والهياكل المؤسسية الجديدة التي تسعى لتأمين المبادئ التوجيهية والسياسـات المشـتركة فيما يتعلق بأنماط تنسـيق البيانات، ونمـاذج هيكلتها، والبيانات الوصفية عنها، وحقوق الملكية الفكرية، وتراخيص الاستخدام القانوني لها، ومعايير المشـاركة بها. وتُشـكل هذه المبادرات مجتمعة مجموعة من مجاميـع البيانات الجديدة - مزيج من نظم الفكر، وأشـكال المعرفة، والتمويل، والاقتصادات السياسـية، والتشريعات الحاكمـة والجوانب القانونيـة والمادية الناظمة، والبني التحتية، والممارسـات، والمنظمات والمؤسسات، والمجتمعات المحلية والشخصيات الاعتبارية، والأماكن، والأسواق - التي تؤطر لكيفية إنتاج البيانات والحدود القصوى لاستخداماتها الممكنة.

ويظهر أثـر البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات بشـكل جلي في مجـالات العلوم، وقطاع الأعـمال، والقطاع الحكومي، والمجتمع المـدني. فالتخصصات الجديدة التي تم اسـتحداثها بـدأت الآن بالتعامل مع تدفق البيانـات الهائل بعد التعود على العمل في جزر البيانات المتفرقة، والسعي نحو استخراج المعلومات، واستخلاص النتائج من عدد قليل نسـبياً من المشاهدات (H.J. Miller 2010). وقد اقترنت هذه التخصصات



بمجـالات جديدة، مثل علـوم البيانات (Data Science)، والحوسـبة الاجتماعية (Social Computing)، والعلوم الإنسـانية الرقمية (Digital Humanities)، والعلوم الاجتماعية الحاسـوبية (Computational Social Sciences)، والتي تعنى بشـكل واضح ببناء البنى التحتيـة للبيانـات وإيجاد طرق مبتكرة لتحليـل وفهم معنى البيانات عند توسـعها وكبر حجمهـا. ففي قطاع الأعمال، توفر البيانات الكبيرة وسـائل جديـدة لإدارة جميع مناحي أنشـطة الشركة بصورة مرنة وفعالة ويسـتفاد منها لتحقيق أرباح إضافية من خلال تعزيز الإنتاجية، والقدرة التنافسـية، والمعرفة بالسـوق. كما أضحت البيانات نفسها سلعة هامة، يتم شراؤها وبيعها مراراً في السوق العالمية بمليارات الدولارات. وبالنسبة للحكومات، وعلى نطاق واسـع، توفـر البيانات الديناميكيـة المتغيرة رؤى جديدة حـول عملياتها، فضلاً عن إعادة تشكيل وسائل الحكم وتنظيم المجتمع. فمن خلال تفحص قواعد البيانات المفتوحة، يسـتخلص المواطنون والمنظـمات غير الحكومية (NGOs) النتائـج الخاصة بهم، متحدين المنظـمات التجاريـة الربحية والأجنـدات الحكومية، ويقدمون رؤى بديلـة للكيفية التي ينبغي بها تنظيم المجتمع وإدارته.

لقد أثارت هذه الفرص الجديدة طفرة حقيقية ملموسة في ما يمكن تسميته بـ "تعزيز البيانات" (Data Boosterism)؛ من خلال حشـد الدعوات المبشرة بفوائد وآفاق البيانات الكبـيرة، والبيانات المفتوحة، والبيانات الصغيرة الموسـعة. والبعـض من هذه الدعوات له مـا يبرره، وبعضها الآخر مجرد غثّ وتهويل دعائي لا حقيقـة له البتة. في المقابل، أصبحت مصطلحـات: البيانات الكبيرة والبيانات المفتوحة، أنماطاً ثقافية قوية في الانتشـار والتأثير (ميمات -Memes)، وليست مجرد وسيلة لوصف البيانات ولكنها رمز لخطاب أوسع مدى يُسـتخدم في حشـد الدعم لها وانتشـار اعتمادهـا والبدء في تطبيقها. فمـع هذا الترويج والتعزيز أضحى من السـهل الانجراف نحو تبني هذه المصطلحات دون انتقاد أو تمحيص للتغيرات التي تحدث والتي أثار العديد منها مخاوف أخلاقية، ومحاذير سياسـية وقانونية. ورغـم ذلك، فقد كشـف التاريخ فيما مضى عن سـوابق من الابتـكارات المربكة المتعلقة بالمعلومات – على سبيل المثال، التحول الجذري في إنتاج المعرفة في أعقاب اختراع الطابعة. ومـن حيث الواقع، فقد صاحب بداية كل عـصر جديد من العلم تقنيات جديدة أدت إلى زيادة مفرطـة في البيانات وكانت بمثابة شرارة التحول نحو طـرق جديدة لإنتاج، وتنظيم،



وتخزين، وتحليل، وتفسـير البيانات (Darnton 2000). على سـبيل المثال، من ملاحظات (Strasser 2012)، فإن الاستكشـافات في عصر النهضة، والتي تعززت بفعل أدوات علمية أفضل في الملاحة، ورسـم الخرائط، قد أسـفرت عن كميات هائلة من الاكتشافات الجديدة التي أدت إلى وسـائل جديدة للتصنيـف، وتقنيات جديدة في التحليـل والتخزين، وأفكار علمية جديدة.

ونظراً للمرحلة المبكرة نسـبياً في ثورة البيانات الحالية، فإنه ليس من المؤكد تماماً كيف ستتكشف وتسـتقر التحولات الحالية، وما الذي سيكون عليه نطاق عواقب التغيرات التي تحدث من منظور أوسـع. فالواضح هو أن هناك حاجة ملحة لمحاولة الإدراك والشعور بما يحـدث. وهكذا، فإن الهـدف من هذا الكتاب هو تقديم إجـمالي، وتحليل مبدئي ودقيق للبيانات وثورة البيانات الجارية حالياً، فهذا الكتاب يسـعى، من جهة، لبيان شـتى الطرق والوسائل المتعلقة بإعادة تشـكيل إنتاج ومعالجة وتحليل وتبادل البيانات، وما يعنيه هذا لكيفية إنتاج واسـتخدام المعلومات والمعرفة؛ وهو من جهة أخـرى يُمهد للنقاش والتفكير النقدي حول البيانات: طبيعتها، وكيفية تأطيرها من النواحي الفنية والفلسـفية والأخلاقية والاقتصاديـة والتجميعـات التقنية والمؤسسـية المحيطة بها. وبدلاً مـن أن تتحدد فوائد البيانـات الكبيرة، والبيانات المفتوحة، والبنـى التحتية للبيانات من خلال حالة عاطفية، أو نقد لاذع يشـجب نتائج العواقب الأكثر سـلبية لها، يقدم الكتاب تقييـماً نقدياً موضوعياً للتغييرات التي تحدث.

ويسـتند التحليل المقدم في هذا الكتاب على مشـاركة واسـعة النطاق مما سـبقه من مؤلفات في مختلف العلوم، والعلوم الإنسانية، والعلوم الاجتماعية، ومن مطبوعات الثقافة الشـعبية، والإصدارات الصحافية، والدوريات الصناعيـة، وعلى خبرة مباشرة في العمل على أنظمة حفظ أرشـيفية وبنى تحتية ومشـاريع تحليلات لبيانات على نطاق مؤسسي واسع. وينقسـم الكتـاب إلى أحد عشر فصلاً. يقـدم الفصل الأول لمحة عامة ونقـداً فكرياً حول مفهـوم البيانـات وكيفية فهم قواعد البيانـات والبنى التحتية للبيانـات. ويبحث الفصل الثاني في الدور المستمر للبيانات الصغيرة والكيفية التي تمّ بها توسيع نطاقها إلى أرشيفات/ محفوظـات رقمية وبنى تحتية، إلى أن وصلت إلى درجة التداول وبيعها من خلال وسـطاء



البيانـات. كما يناقـش الفصل الثالث الدافـع للتوجه نحو اسـتحداث البيانات المفتوحة والبيانات المرتبطة التي يجري تقاسـمها ومشـاركتها وإعادة اسـتخدامها على نطاق أوسع بمـرور الوقت. ويبيّن الفصلان الرابع والخامس طبيعـة البيانات الكبيرة بالتفصيل وعوامل التمكين للبيانات الكبيرة ومصادرها المسـاعدة على انتشارها، في حين يقدّم الفصل السادس لمحة عامة عن مجموعة جديدة من تحليلات البيانات التي ترمي إلى فهم البيانات الصغيرة الموسعة والبيانات الكبيرة. كما يدرس الفصلان التاليان (السابع والثامن) الحجج المستخدمة لتعزيز البيانات الكبيرة والترويج لها وبيان تأثيرها في الحوكمة والأعمال، والسبل التي تقوم بها ثورة البيانات على إعادة تشـكيل كيفية تصوّر البحوث وممارسـتها. ويناقش الفصلان التاسـع والعاشر التحديات التقنيـة والتنظيمية، والأخلاقية، والسياسـية، والقانونية لثورة البيانـات. ويحدد الفصل الأخير بعض الاسـتنتاجات الكلية ويقـدّم خارطة طريق لإجراء المزيد من البحث والتأصيل.




المكتبة الاقتصادية
ECONLIBRARY
قناة المكتبة على التليجرام





الفصل الأول

تصور لمفهوم البيانات (Conceptualising Data)

يتمّ فهم البيانات عادة على أنها المادة الخام التي تنتجها طرق تجريد العوالم من حولنا والتعبـير عنها مـن خلال فئات، ومقاييس، وأشـكال تمثيلية أخـرى - كالأرقام، والحروف، والرموز، والصور، والأصوات، والموجات الكهرومغناطيسية، والثنائية الرقمية - تشكل اللبنات الأساسية في بناء صروح المعلومات والمعرفة. ومع أن البيانات تُعد بطبيعتها مُعبرة ومُمثلة للأشـياء (على سبيل المثال، القياسات المتعلقة بظاهرة ما كعمر الشخص، طوله، وزنه، لونه، ضغـط دمه، رأيه، عاداته، مـكان إقامته، وما إلى ذلك)، إلا أنه مـن الممكن أن تكون أيضاً ضمنية (من خلال الغياب بدلاً من الوجود)، أو مُشتقة على سبيل المثال، البيانات التي يتم إنتاجها من البيانات الأخرى كالنسـبة المئوية للتغير مع مرور الوقت والتي يتم احتسـابها مـن خلال مقارنة البيانات من فترتين زمنيتين مختلفتين، ويمكن أن تكون البيانات مُسـجلة ومُخزنة إما في شـكل تناظري (Analog) أو من خلال ترميزها في شكل رقمي كأرقام ثنائية (Bits). وتتصـف البيانـات بأنها ذات جودة جيدة إذا كانـت منفصلة وواضحة بمعنى أن تكون كل وحدة بيان (Datum) مفردة، ومستقلة وقابلة للانفصال، ومعرفة بشكل واضح، وتراكمية قابلة للحصر (يمكن أن تبنى في مجموعات)، ومرتبطة ببيانات وصفية لها (بيانات عـن البيانات)، ويمكن ربطها بمجموعات بيانات أخرى لتقديم رؤى وأفكار غير متوفرة من مجموعـة بيانات واحدة (Rosenberg 2013). وتتمتع البيانات باسـتخدام قوي وقيمة عالية كونها توفر المدخلات الأساسـية لكافة مراحـل التحليل التي يجري توظيفها من قبل الأفراد، والمؤسسات، وقطاعات العلوم والأعمال لفهم وتفسير العالم الذي نعيش فيه، والتي تُستخدم بدورها لاستحداث الابتكارات، المنتجات، السياسات، والمعارف التي تشكل الكيفية التي يعيش بها الناس حياتهم.

ومـن هنا، تعد البيانات مورداً رئيسـياً في العصر الحديث. ومع ذلـك، ونظراً لفائدتها وقيمتها، ومقدار الجهد والمـوارد المخصصة لإنتاجها وتحليلها، فمن اللافت للنظر كيف تم إيـلاء القليل من الاهتمام للتصـور المفاهيمي والإدراك المبدئي للبيانـات في حدَ ذاتها، وفي

المقابل، هناك الآلاف من المقالات والكتب المكرسة لفلسفة المعلومات والمعرفة. فكما أننا نميل في العادة للتركيز على المباني والأحياء عندما ننظر إلى المُدن، بدلاً من الطوب والإسمنت المستخدم في بنائها، فكذلك هو الحال مع البيانات. وعلاوة على ذلك، وبالمثل، كما اعتدنا التفكير بالطوب والإسمنت بصفتها مجرد وحدات بنيوية بدلاً من عناصر تم تصنيعها داخل مصانع من قبل شركات ملزمة ضمن اعتبارات لوجستية ومالية وقانونية، وأنه يتم توزيعها، وتخزينها، وتداولها، كذلك حالنا إلى حد كبير في التعامل مع البيانات. ونتيجة لذلك، وعندما تكون البيانات هي محور التساؤل والاهتمام، عادة ما يتم النظر، بالمعنى التقني إلى حد كبير، إلى الكيفية التي ينبغي بها إنتاج وتحليل البيانات، أو الكيفية التي يمكن بها الاستفادة منها في إنتاج رؤى وقيم، وليس زيادة النظر في طبيعة البيانات من منظور مفاهيمي وفلسفي.

ومع الأخذ بهذه الملاحظة في الاعتبار، فإن الهدف الرئيسي من هذا الكتاب هو مركب ثلاثي الأبعاد، فيه: توفير شرح تفصيلي عن طبيعة البيانات ومجاميعها على نطاق أوسع؛ وإيضاح الكيفية التي تتحول فيها هذه المجاميع وتتحور مع تطوير بنى تحتية جديدة للبيانات، والبيانات المفتوحة، والبيانات الكبيرة؛ والتفكير المعمق في الآثار المترتبة عن هذه المجاميع الجديدة للبيانات فيما يتعلق بكيفية استشعار العالم المحيط بنا والتعاطي معه. ولتوفير منهاج مفاهيمي وتصور مبدئي أولي، سيتم في هذا الفصل فحص أنماط البيانات وأطرها، وطبيعتها، والقواعد الفلسفية لها بالتفصيل. وبعيداً عن كونها وحدات بنيوية بسيطة، تكشف المناقشة عن أن البيانات أكثر تعقيداً مما كنا نظن. ففي حين يقبل العديد من المحللين البيانات على ظاهرها، ويعاملونها كما لو أنها محايدة وموضوعية وقابلة مسبقاً للتخمين والتحليل في طبيعتها، إلا أن البيانات مؤطرة في الواقع من الناحية الفنية، والأخلاقية، والاقتصادية، والزمانية، والمكانية، والفلسفية. فالبيانات لا يمكن أن توجد بشكل مستقل عن الأفكار، والأدوات، والممارسات، والسياقات، والمعارف المستخدمة لإنتاجها، ومعالجتها، وتحليلها (Bowker 2005; Gitelman and Jackson 2013). ولذا، فإن الحُجة المُقدمة هي أن فهم البيانات وثورة البيانات الجارية حالياً يتطلب تحليلاً أكثر دقة من ذلك الذي تستعرضه الكثير من الأدبيات ذات الصلة بالبيانات المفتوحة والبيانات الكبيرة في وقتنا الحاضر.



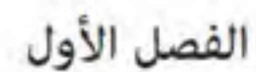
ماهية البيانات:

بشكل اشـتقاقي، فإن كلمة البيانات «Data» مشـتقة من الكلمة اللاتينية «Dare»، التي تعني «أن تُعطي». وبهذا المعنى، فإن البيانات هي عناصر خام يمكن استخراجها من أو تُعطى بواسـطة ظاهرة معينة يجري قياسـها وتسجيلها بطرق مختلفة. ولكن البيانات تُشـير في الاسـتخدام العام إلى تلك العناصر التي تُؤخذ وتُسـتخرج من خلال الملاحظات، والحسـابات، والتجارب، وحفظ السـجلات (Borgman 2007). أما مـن الناحية الفنية، فما نفهمه مـن بيانات أنها في الواقع التقاطات «Capta» والمشـتقة من الكلمة اللاتينية «Capere»، والتـي تعنـي «الأخـذ»؛ أي تلك الوحدات مـن البيانات التي تـم اختيارها وانتقاؤهـا من مجموع كل البيانات المحتملـة (Kitchin and Dodge 2011). كما قال (1950: ix, مقتبس من Jensen (Becker 1952: 278 :

> كان حادثاً مؤسـفاً في التاريخ أن يستخدم المصطلح بيان «Datum» بدلاً من مصطلـح لقطة «Captum» للدلالة على وصف وحدة الظاهرة في العلوم. وهي ليست المناولات العلمية التي منحتها الطبيعة للعالم، ولكن المناولات التي أُخذت أو انتقيت من الطبيعة من قبل العالم نفسه وفقاً لهدفه.

من ثـم، وبالمعنى الدقيق للكلمـة، ينبغي على الكتاب أن يُعنـون بثورة اللقطة «The Capta Revolution». ومـع ذلك، ولأن مصطلح البيانات «Data» قد أصبح متأصلاً في لغة الاسـتخدام الأكاديمي وقطاع الأعـمال لتعني اللقطة «Capta»، وبدلاً من الخلط واللغط أكثر في هذه المسألة فمن المنطقي المتابعة والاستمرار في استخدام مصطلـح البيانات «Data» مع أن اعتماد مصطلح اللقطـة «Capta» أكثر ملائمة في الاستخدام. وبعد تجاوز تسـليط الضوء على جذور الاشتقاق اللغوي لهذا المصطلح، فإن هذه المناقشـة الوجيزة تبدأ في تسـليط الضوء على أن البيانات التي يتم جمعها مـن خلال القياس هي دائمـاً لمجموعة منتقاة من مجمـوع كل البيانات المحتملة - والتي اخترنـا التقاطها من كل المعطيات الممكنة. وعلى هـذا النحو، تكون البيانات جزئية بطبيعتها، انتقائية، وممثلة، مع أهمية المعايير المميزة المستخدمة في التقاطها.





وقـد لاحظ علماء آخرون أن ما تم فهمه على أنــه «البيانات» قد تغير مع مرور الزمن وتطـور العلم. فقد بيّن (Rosenberg 2013) بأن مصطلح «البيانات» اسـتخدم لأول مرة باللغـة الإنجليزية في القرن السـابع عشر. وارتبط كمفهوم إلى حد كبـير بالحداثة والنمو والتطور المصاحب للعلوم، والطرق الجديدة في إنتاجٍ، وعرض، ومناقشـة المعرفة في القرنين السـابع عشر والثامـن عشر والتي انتقلت بعيـداً عن اللاهوت، والوعـظ، والعواطف إلى الحقائـق والأدلة واختبار النظرية مـن خلال التجربـة (Poovey 1998; Garvey 2013; Rosenberg 2013). ومـع مـرور الوقت، جاء فهـم البيانات على أنها سـابقة للتحليل، وهـي ما قبل اتخاذ صفة الواقعية، وتختلف في طبيعتها عن الحقائق، والأدلة، والمعلومات، والمعرفة، ولكنها عنصر مؤسـس في تكوين جميع هذه العناصر على الرغم، في الغالب، من اختلاط التعاريف والمصطلحات المسـتخدمة في وصف البيانـات، الحقائق، الأدلة، المعرفة، والمعلومات بحسب ما أشار إليه (Rosenberg 2013: 18) حين ذكر أن:

> الحقائـق وجوديـة (Ontological)، والأدلـة معرفيـة (Epistemology)، والبيانات بلاغيـة (Rhetorical). وتكون وحدة البيان «Datum» حقيقة أيضاً، كما قد تكون الحقيقة في واقع الأمر دليلاً.... على أن وجود وحدة البيان «Datum» يُعـد مسـتقلاً دون أي اعتبار للحقيقـة الوجودية المقابلـة (Corresponding Ontological Truth). وعنـد نفي الحقيقة وإثبات عدم صحتها، فإنها تتوقف عن كونها حقيقة. فالبيانات الكاذبة هي بيانات مع ذلك.

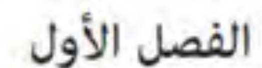
وبحسـب المصطلحات/التعابير البلاغية، فالبيانـات هي تلك التي توجد قبل النقاش أو التفسير الذي يحولها إلى حقائق، وأدلة، ومعلومات (Rosenberg 2013). ومن هذا المنظور، تحمل البيانات مفاهيم مميزة: فهي تكون بذلك مجردة، ومنفصلة، وتراكمية (يمكن الجمع بينها) (Rosenberg 2013)، وذات معنى مستقل عن الشكل، واللغة، والمُنتج لها، والسياق (بمعنى، أن تحمل البيانات معناها سـواء تم تخزينها بصورة تناظرية (Analog) أو رقمية (Digital)، أو تم عرضها على الورق أو الشاشة أو تم التعبير عنها بأية لغة، أو تم تقييدها بأنماط محددة غير متغيرة، مثل إشارة عدد الحلقات في جذوع الأشجار لعمر الشجرة والذي يكون دوماً مساوياً له) (Floridi 2010). ويؤكد (Floridi 2008) أن دعم استقلال البيانات



قائم على ثلاثة أنواع من الحيادية هي: التصنيفية (Taxonmic) كون البيانات تُمثل كيانات علائقية معرفة تحديداً بالنسبة إلى بيانات خاصة أخرى؛ والطوبوغرافية (Typological) لأنه يمكن للبيانات اتخاذ أشكال مختلفة منفصلة بعضها عن بعض، على سبيل المثال، الشكل الأساسي، الثانوي، الوصفي، التشغيلي، والمشتق؛ والجينية (Genetic) كون البيانات تمتلك دلالات مستقلة عن مدى فهمها، فعلى سبيل المثال، تشكل الهيروغليفية على حجر الرشيد بيانات بغض النظر عن حقيقة أنه عند اكتشافه لم يمكن لأحد أن يفسرها.

وفي واقع الأمر ليس كل من يفكر أو يشتغل بالبيانات يحمل هذه النظرة البلاغية الضيقة. ففهم البيانات لم يتطور مع مرور الوقت فحسب، بل اختلف هذا الفهم بحسب منظور الفهم ووجهة النظر للبيانات. على سبيل المثال، أوضح (Floridi 2008) أن البيانات من المنظور المعرفي هي مجموعات من الحقائق، ومن المنظور المعلوماتي البيانات هي معلومات، ومن منظور الحوسبة التقنية البيانات هي مجموعات من العناصر الثنائية التي يمكن معالجتها ونقلها إلكترونياً، ومن المنظور العام (Diaphoric) البيانات هي عناصر مجردة بتميز واضح وملموس من البيانات الأخرى، ففي الحالة الأولى، توفر البيانات الأساس لمزيد من الاستدلال أو تشكل الأدلة التجريبية. وفي المجموعة الثانية، تشكل البيانات معلومات تمثيلية/ممثلة يمكن تخزينها ومعالجتها وتحليلها، ولكن لا تشكل بالضرورة حقائق. أما في المجموعة الثالثة، فتشكل البيانات المدخلات والمخرجات من الحوسبة ولكن ينبغي معالجتها لتحويلها إلى حقائق ومعلومات على سبيل المثال، يحتوي القرص الرقمي المدمج (DVD) على غيغابايت من البيانات، وهي ليست وقائع أو معلومات بحد ذاتها (فلوريدي، ٢٠٠٥). في المنظور الرابع، تكون البيانات ذات مغزى لأنها التقاط ودلالة على التنوع والتغيير على سبيل المثال، أنماط من النقاط، والحروف الأبجدية والأرقام، والموجات الطولية التي تقدم إشارة يمكن تفسيرها. وكما هو مبين أدناه، قد يشمل فهم البيانات مناظير أخرى مثل الفهم القائم على اعتبار البيانات بناء اجتماعياً، أو امتلاك البيانات وجوداً نسبياً، أو تحميلها مضامين إيديولوجية، أو اعتبارها سلعة يتم تداولها الاتّجار بها، أو أنها تشكل منفعة عامة، وغير ذلك من المفاهيم الموجودة. والنقطة الأساسية هنا، أن البيانات لم تكن أبداً ببساطة مجرد بيانات. حيث تختلف الكيفية التي يتم فيها تصور البيانات واستخدامها من قبل أولئك الذين يجمعون، ويحلّلون، ويستخلصون النتائج منها.





أنواع البيانات:

سـواء كانت البيانات بطبيعتها قابلة للتخمين (تم التكهن قبل التحقق منها) ومطابقة بلاغيـاً أو لا، فمن الواضـح أن البيانات متنوعة في خصائصها، مما يحـدّد بعبارات واضحة كيفية التعامل معها وما يمكن القيام به من خلالها. وبصفة عامة، تختلف البيانات بحسب النمـوذج (النوعي أم الكمّي)، والهيكل (المنظم، أو شـبه المنظم، أو الغير منظم)، والمصدر (مسـجّلة، أم مشتقة، أم مسـتنفذة، أم عابرة)، والمُنتج (رئيسي، أو ثانوي، أو عام)، والنوع (بيانات تأشيرية، أو بيانات خاصية، أو بيانات وصفية).

البيانات الكمية (Quantitative Data) والبيانات النوعية (Qualitative Data):

من الممكن أن تتخذ البيانات العديد من الأشكال المادية متضمناً ذلك الأرقام، والنصوص، والرموز، والصور، والصوت، والموجات الكهرومغناطيسية، أو حتى الفراغ أو الصمت (فالفضاء الفـارغ هو في حدّ ذاته بيانات). وتنقسـم هـذه البيانات في العادة إلى فئتين رئيسـيتين: البيانات الكمية والبيانات النوعية، فالبيانات الكمية تتكون من سجلات رقمية، وبشكل عام تكون هذه البيانات واسـعة النطاق وتتعلق بالخصائـص الفيزيائية للظواهر (مثل الطول، الارتفاع، المسافة، الوزن، المساحة، والحجم)، أو تكون تمثيلية وتتعلق بالخصائص غير المادية للظواهـر (مثل الطبقة الاجتماعية، المسـتوى التعليمي، الحرمـان الاجتماعي، وتصنيفات مستوى المعيشـة). وللبيانات الكمية أربعة مستويات مختلفة للقياس والتي تحدد كيفية معالجتهـا وتحليلها (Kitchin and Tate 1999، وانظر أيضاً الجدول ١-١). ويمكن تحليل هذه البيانات باسـتخدام العروض المرئية، ومجموعة متنوعـة من طرق الإحصاء الوصفي والاستدلالي، وتستخدم باعتبارها مدخلات للنماذج التنبؤية ونماذج المحاكاة.



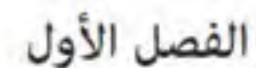
الجدول (١-١)

مستويات قياس البيانات

مستويات القياس	التعريف	المثال
البيانات الاسمية (Nominal data)	فئوية بطبيعتها، مع تسجيل المشاهدات في وحدات قياس منفصلة.	غير متزوج، متزوج، مطلّق، أرمل
البيانات النظامية الترتيبية (Ordinal data)	ترتب المشاهدات في فئات منظمة، حيث بعض المشاهدات هي أكبر من غيرها.	منخفضة، متوسطة، عالية.
بيانات الفترة (أو المسافة) (Interval data)	تمتد القياسات على طول مقياس يتضمن الفترة أو المسافة بين نقطة أصل ثابته وأخرى متغيرة. ولا تتغير طبيعة فترة المشاهدات في هذا النوع من القياسات عند الإضافة أو المضاعفة بمقدار ثابت. ويمكن أن تكون قيمة البيانات متصلة (على سبيل المثال، الوقت أو الطول) أو منفصلة/متقطعة (على سبيل المثال، عدد مرات المشاهدة) في الطبيعة.	درجة الحرارة على المقياس المئوي (Celsius scale).
البيانات النسبية (Ratio data)	مماثلة لبيانات الفترة عدا أنها ذات نقطة أصل صفرية حقيقية، كما أن المضاعفة بمقدار ثابت لن تغير الطبيعة النسبية للمشاهدة.	علامة الاختبار على مقياس من صفر-١٠٠

وفي المقابل، فإن البيانات النوعية هي بيانات غير عددية، مثل النصوص، الصور، الفن، الفيديو، الأصوات، والموسيقا. وبرغم أنه يمكن تحويل البيانات النوعية إلى بيانات كمية، إلا أن ذلك يتضمن خفضاً وتجريداً كبيرين وخسارة لثراء البيانات الأصلية نتيجة لعملية التحويل. ولذا، فإن عملية تحليل البيانات النوعية تُمارس في العادة على المواد الأصلية، سعياً لاستخلاص وبناء المعنى والفهم بدلاً من إخضاع البيانات ومجرد إمرارها لتقنيات حاسوبية. ومع ذلك، فقد تم إحراز تقدّم كبير فيما يتعلق بمعالجة وتحليل البيانات النوعية حاسوبياً من خلال عدد من التقنيات مثل التعلّم الآلي والتنقيب في البيانات (انظر الفصل السادس).





البيانات المنظمة (Structured Data)، والبيانات غير المنظمة (Unstructured Data)، والبيانات الشبه المنظمة (Semi-structured Data):

البيانات المنظمة هي تلك البيانات التي يمكن تنظيمها، وتخزينها، ونقلها بسهولة من خلال نموذج بيانات محدد، مثل قائمة الأرقام والنصوص الواردة في جدول أو قاعدة بيانات علائقية ذات تنسيق ثابت موحّد (على سبيل المثال، الاسم، تاريخ الميلاد، العنوان، الجنس،... إلخ). ويمكن معالجة هذه البيانات، والبحث فيها، والاستعلام عنها، والجمع بينها، وتحليلها بشكل مباشر نسبياً باستخدام حساب التفاضل والتكامل والخوارزميات، كما يمكن تمثيلها صورياً باستخدام أشكال مختلفة من الرسومات البيانية والخرائط، ومعالجتها بسهولة من قبل أجهزة الحاسب الآلي. أما البيانات الشبه المنظمة فهي بيانات ذات تنظيم/ هيكل فضفاض وليس لها نموذج أو مخطط مسبق، ومن ثم لا يمكن تقييدها ضمن قاعدة بيانات علائقية. وتكون بنية البيانات الشبه المنظمة غير متسقة، وضمنية، ومرنة، وغالباً ما تتداخل في شبكة ذات شكل هرمي، ولكنها في ذات الوقت تتمتع بمجموعة متسقة بشكل كبير من حقول البيانات، ويتم تمييز كل منها بوسم مميز، ومن ثم يُمكن فصل المحتوى بصورة دلالية وتوفير بيانات وصفية ذاتية التعريف بصورة مرنة واستخدامها وسيلة لفرز وترتيب وهيكلة البيانات. ومن الأمثلة على البيانات شبه المنظمة صفحات الويب الموسومة بلغة التوصيف الموسعة (Extensible Markup Language- XML) وهي الصفحات التي تستخدم لغة التوصيف الموسعة (XML) لتنسيق وترميز المستندات بصورة قابلة للقراءة آلياً وبشرياً في آن واحد؛ Franks 2012؛ انظر البيانات المرتبطة في الفصل الثالث).

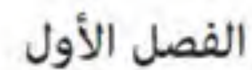
في المقابل، لا تمتلك البيانات غير المنظمة نموذج بيانات محدّداً أو هيكلاً تعريفياً مشتركاً. ولكن قد يكون لكل عنصر على حدة شكل أو هيكل محدّد خاص به، مثل عناصر النص الروائي أو الصورة، ولكن ليس كل البيانات ضمن مجموعة البيانات لها الهيكل نفسه. على هذا النحو، وبينما يمكن في الغالب البحث والاستعلام عن البيانات غير المنظمة، إلا أنه ليس من السهل الجمع بينها وتحليلها باستخدام الحاسب الآلي. وتكون البيانات غير المنظمة نوعية بطبيعتها، ولكن من الممكن في أغلب الأحيان تحويلها إلى بيانات منظمة من خلال الفرز والتصنيف. وحتى وقت قريب نسبياً، كانت مجموعات البيانات الكبيرة

جـداً تحوي في العموم بيانات منظمة كونها أسـهل بكثير في المعالجـة والتحليل والتخزين على هذه الحال. أما في عصر البيانات الكبيرة، فقد تكونت العديد من مجموعات البيانات الضخمـة التي تحوي بيانات شـبه منظمة أو غـير منظمة، مثل التي أنتجتها المشـاركات الاجتماعية الجماعية على مواقع التواصل الاجتماعي مثل الفيسبوك، التغريدات على تويتر، الصور ومقاطع الفيديو المرفوعة على الشـبكة العالمية، والمدونات (Blogs)، إذ تشير بعض التوقعات إلى معدل زيادة في نمو هذه البيانات غير المنظمة أو الشبه المنظمة بمقدار خمسة عـشر ضعفاً عن البيانات المنظمـة (Zikopoulos et al. 2012)، وقد صاحب ذلك تطورٌ في تصميـم قواعد البيانات مثل قواعد البيانات غير العلاقية (NoSQL) والتي لا تسـتخدم النـماذج الجدولية أو قواعد البيانات العلائقية، انظر الجدول رقم ٥)، وتقنيات التعلم الآلي التي تساعد في التخزين والتحليل (انظر الفصل السادس).

البيانـات الملتقطة (Captured Data)، والبيانات المسـتنفدة/ المسـتهلكة (Exhaust Data)، والبيانات العابرة (Transient Data)، والبيانات المشتقة/ المستمدة (Derived Data):

هنـاك طريقتان رئيسـيتان لإنتاج البيانات: الأولى هي أنه مـن الممكن الحصول على البيانات بصورة مباشرة من خلال شـكل من أشكال القياس مثل الملاحظة، والمسح الميداني، والتجارب المخبرية والميدانية، وحفظ السجلات (على سبيل المثال، ملء الاستمارات أو كتابة المدونات)، والكاميرات، والماسـحات الضوئية، وأجهزة الاستشعار. وفي هذه الحالات، عادة مـا تكون البيانات هي المُنتج المطلـوب للقياس؛ أي أن النية هي إنتاج بيانات مفيدة. وفي المقابل، يتم إنتاج البيانات المسـتنفدة بالأصل بواسطة جهاز أو نظام، وتكون ناتج وظيفة رئيسـية وليست المخرج الرئيسي (Manyika et al. 2011). فعلى سبيل المثال، فإن جهاز تسـجيل الخروج الإلكتروني المصمم لاحتسـاب إجمالي البضائع التي تـّم شراؤها ومعالجة عملية الدفع لها من الممكن أن يُنتج كذلك بيانات تستخدم لمراقبة المخزون، وأداء العاملين، وعمليات شراء العملاء. وتُنتج العديد من الأنظمة المحوسبة مثل هذه البيانات المستنفدة والتي أصبح الكثير منها مصدراً قيماً للبيانات. وفي حالات أخرى، فإن البيانات المسـتنفدة تبقى بيانات عابرة بطبيعتها؛ أي أنه لم يسبق أن جرى فحصها أو معالجتها ولذا فإنه يجري





التخلص منها، إما لأنها ضخمة جداً أو غير منظمة بطبيعتها، أو مكلفة في المعالجة والتخزين، أو أنه يوجد نقص في تقنيات استخلاص القيمة منها، أو لقلة الاستخدام الإستراتيجي أو التكتيكي لها (Zikopoulos et al. 2012; Franks 2012). ومن الأمثلة على ذلك ما ذكره Manyika et al. (2011: 3) من أن "مقدمي خدمات الرعاية الصحية.... يتجاهلون ٩٠٪ من البيانات التي ينتجونها (كالفيديوهات المسجلة أثناء العمليات الجراحية)".

وتُعد البيانات الملتقطة والبيانات المستنفدة بيانات «خام» بمعنى أنه لم يتمّ تحويلها أو ضمّها مع غيرها من البيانات. وفي المقابل، يتم إنتاج البيانات المشتقة من خلال عمليات إضافية من المعالجة والتحليل للبيانات التي تمّ التقاطها. فعلى سبيل المثال، قد تكون البيانات هي تعداد حركة الأفراد على أحد التقاطعات المرورية والبيانات المشتقة منها هي التعداد الإجمالي أو التعداد لكل ساعة، فالبيانات الأخيرة استمدت من البيانات السابقة. وتكون البيانات التي جرى التقاطها في كثير من الأحيان مدخلاً إلى نموذج العملية، في حين تكون البيانات المشتقة هي المخرج عن نموذج العملية. فعلى سبيل المثال، قد تكون بيانات الكثافة المرورية مدخلاً إلى نموذج النقل والمواصلات على أن يكون المخرج هو بيانات التوقع أو المحاكاة (مثل حساب حركة المرور المتوقعة في أوقات مختلفة أو في ظل ظروف مختلفة). ففي حال وجود نموذج، من المرجح أن تضم بيانات التعداد/ الكثافة المرورية مع بيانات أخرى ملتقطة أو مشتقة (مثل نوع السيارة، وعدد الركاب،.. إلخ) لإنشاء بيانات مشتقة جديدة قد تشكل مدخلاً للنموذج. ويتمّ إنتاج البيانات المشتقة لعدة أسباب، بما فيها الحدّ من حجم البيانات والتقليل منها إلى قدر مقبول من الممكن إدارته لإنتاج مقاييس أكثر فائدة وذات معنى. وفي بعض الأحيان، قد تتمّ معالجة البيانات الأصلية الملتقطة إلى مستويات مختلفة من الاشتقاق اعتماداً على الاستخدام المقصود. ويُعد نظام وكالة الفضاء الأمريكية - ناسا الخاص برصد ومراقبة الأرض من الأمثلة على ذلك حيث يُقدم بياناته في ستة مستويات تبدأ من البيانات الملتقطة وغير المعالجة، وتمضي خلال درجات متزايدة من مستويات المعالجة والتحليل وصولاً إلى مخرجات النموذج القائمة على تحليل بيانات المستوى الأقل (Borgman 2007؛ انظر جدول ٢-١).

الجدول (٢-١)

المستويات الستة لبيانات نظام رصد ومراقبة الأرض الخاص بوكالة الفضاء الأمريكية

مستوى البيانات	الوصف
المستوى	بيانات غير منظمة، وغير معالجة بالدقة الكاملة وبالحالة الخام وتحوي كافة إضافات بيانات الاتصالات (على سبيل المثال، بيانات التزامن، ترويسة الاتصال، البيانات المكررة)
المستوى الأول أ	بيانات غير منظمة، وغير معالجة بالدقة الكاملة، ولكن معنونة زمانياً ومذيّلة أو متبوعة بشروحات إضافية، تشمل معاملات معايرة ومقاييس راديوية وجغرافية ومؤشرات تحتسب وتضاف للبيانات في المستوى ٠ ولكن لا تطبق عليها.
المستوى الأول ب	هي بيانات المستوى الأول أ ولكن تمّ معالجتها إلى وحدات الاستشعار
المستوى الثاني	متغيرات جيوفيزيائية مشتقة بنفس دقة وموقع مصدر بيانات المستوى الأول
المستوى الثالث	متغيرات تمّ تعيينها في مقاييس شبكة منتظمة للفراغ - الزمن، في العادة مع بعض المكملات والتدقيق على سلامتها
المستوى الرابع	مخرج النموذج أو النتائج من التحليل لبيانات المستويات الأقل (المتغيرات المشتقة من قياسات متعددة).

المصدر: مقتبس من الموقع الإلكتروني على العنوان :

https://earthdata.nasa.gov/data/standards-and-references/processing-levels

البيانات الرئيســية (Primary Data)، البيانات الثانوية (Secondary Data)، والبيانات العامة (Tertiary Data):

يتم إنتاج البيانات الرئيســية (الأساســية) من قبل الباحثين وأدواتهم من خلال تصاميم البحــوث التــي يضعونها. أما البيانــات الثانوية فهي البيانات التي يتم إنشــاؤها من قبل شــخص ثمَ تتاح للآخرين لإعادة استخدامها وتحليلها. وعلى ذلك، يمكن أن تكون البيانات الأساسية لشخص ما هي بيانات ثانوية عند شخص آخر. أما البيانات العامة فهي شكل من أشكال البيانات المشتقة، مثل التعداد، والفئات، والنتائج الإحصائية، وتصدر البيانات العامة غالبــاً من قبــل الوكالات الإحصائية بدلاً مــن البيانات الثانوية لضمان السرية للأشــخاص الذين تشــير إليهم هذه البيانات. على ســبيل المثال، يحظر نشر البيانات الأساسية للتعداد الســكاني الأيرلندي على أنها بيانات ثانوية إلا بعد مائة عام من إنتاجها؛ ولكن يتم الإفراج عــن البيانــات على صورة ملخصات إجمالية وبيانات تصنيفية عامة. ويســعى العديد من الباحثين والمؤسســات لإنتاج البيانات الأساسية لأنها مصممة خصيصاً لاحتياجاتهم الخاصة ومركّــز عليها، في حــين لا تتوفر خيارات التصميم هذه لأولئك الذيــن يقومون على تحليل البيانــات الثانويــة أو العامة. وعلاوة على ذلــك، على أولئك الذين يســتخدمون البيانات الثانويــة والعامة كمدخلات في دراســاتهم الخاصة أن يتوثقوا مــن صلاحية البحث الأصلي وصحته.

ويجمع الباحثون في كثير من الأحيان بين البيانات الأساســية والبيانات الثانوية والعامة لإنتاج بيانات مشــتقة ذات قيمة أكبر. على ســبيل المثال، قد تسعى متاجر التجزئة لإنشاء مجموعة بيانات مشــتقة تدمج بيانات المبيعات الأساسية مع البيانات السكانية الجغرافية العامــة (Tertiary geodemographics data)، وهي بيانات عن أي نوع من الناس الذين يعيشــون في مناطق مختلفة والمســتمدة من التعداد الســكاني والبيانات العامة والتجارية الأخرى من أجل تحديد أماكن لاســتهدافها بالمواد التسويقية. إن البيانات الثانوية والعامة ذات قيمة لأنها تُمكن من إجراء الدراســات التكرارية وبنــاء مجموعات بيانات أكبر حجماً وأكثر ثراء وتعقيداً. وهي تنتج لاحقاً ما أطلق عليه (Crampton et al. 2012) مصطلح تضخيم البيانــات (Data Amplification) والذي يعني أن الجمع بين البيانات يتيح رؤى

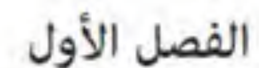


أكبر بكثير من خلال الكشف عن الروابط، العلاقات، والأنماط التي تبقى خافية في حال ظلت البيانات معزولة، ونتيجة لذلك، فإن بيانات السوق الثانوية والعامة هي صناعة بمليارات الدولارات (انظر الفصل ٢).

البيانات التأشيرية (Indexical Data)، البيانات الخاصية (Attribute Data)، والبيانات الوصفية (Metadata):

تختلف البيانات أيضاً في النوع، فالبيانات التأشيرية هي تلك البيانات التي تتيح التعريف بالبيانات وربطها بغيرها من البيانات الأخرى، وتشمل معرّفات فريدة من نوعها، مثل أرقام جوازات السفر وبطاقات الضمان الاجتماعي، وأرقام بطاقات الائتمان، والأرقام التسلسلية المصنعة، ومعرفات الكائن الرقمي، وعناوين الربط الشبكي (IP and MAC)، وأرقام الشحن وطلبات الشراء، إضافة إلى الأسماء، والعناوين، والرموز البريدية. إن البيانات التأشيرية مهمة لأنها تساعد في حصر كميات كبيرة من البيانات غير المفهرسة بعضها مع بعض وتتبعها من خلال معرّفات مشتركة، والمساعدة على التمييز بينها، وضمّها، وتقسيمها وإعادة تجميعها، والبحث فيها وغيرها من أشكال المعالجة والتحليل. وكما هو مبين في الفصل الرابع، فقد أضحت البيانات التأشيرية شائعة على نحو متزايد وبصورة متخصصة، مما يزيد من درجة الارتباط والعلائقية بين مجموعات البيانات. أما البيانات الخاصية فهي بيانات تمثل جوانب من الظاهرة، ولكنها ليست تأشيرية بطبيعتها، وعلى سبيل المثال، بالنسبة إلى شخص ما، قد تكون البيانات التأشيرية هي بصمات الأصابع أو تسلسل الحمض النووي (DNA)، وقد تكون بيانات الخاصية له مثل العمر، الجنس، الطول، الوزن، لون العينين، فصيلة الدم، وغيرها. مع ملاحظة أن الغالبية العظمى من البيانات التي يتم إنشاؤها وتخزينها في الأنظمة هي بيانات الخاصية.

البيانات الوصفية هي بيانات عن البيانات، ويمكن أن تشير البيانات الوصفية إلى محتوى البيانات أو مجموعة البيانات كاملة، وقد تشمل البيانات الوصفية عن المحتوى أسماء ووصف حقول بيانات محددة على سبيل المثال، رؤوس الأعمدة في جدول البيانات وتعريفات البيانات. وتساعد هذه البيانات الوصفية مستخدم مجموعة البيانات على فهم مكونات البيانات وتركيبتها وكيف ينبغي استخدامها وتفسيرها، وتساعد في الدمج بين

مجموعات البيانات، وقابلية التشغيل البيني وتبادل البيانات بينها، والاكتشاف، والقدرة على الحكم على مصدر البيانات وسندها. أما البيانات الوصفية التي تشير إلى مجموعة البيانات كلها فلها ثلاثة أشكال (NISO 2004). فالشكل الأول يتعلق بالبيانات الوصفية التوضيحية والتي تشير إلى بيانات التحديد/التعريف والاكتشاف المتعلقة بمجموعة البيانات، وتشمل عناصر مثل الاسم، والمؤلف، والناشر، والموضوع، والوصف. ويتضمن الشكل الثاني البيانات الوصفية الهيكلية والتي تشير إلى بيانات التنظيم والتغطية لمجموعة البيانات. ويحتوي الشكل الثالث البيانات الوصفية الإدارية التي تعكس متى وكيف تم إنشاء مجموعة البيانات، وتفاصيل الجوانب الفنية للبيانات، مثل صيغة تنسيق الملف، ومن الذي يملك وله صلاحية استخدام البيانات. ومن المعايير الشائعة للبيانات الوصفية والذي يجمع الأنواع الثلاثة للبيانات الوصفية معيار دبلن كور (Dublin Core) (http://dublincore.org). ويشترط هذا المعيار تمتع مجموعة البيانات بخمسة عشر حقلاً للبيانات الوصفية هي : العنوان، المنشىء، الموضوع، الوصف، الناشر، المساهم، التاريخ، النوع، الصيغة، المعرّف، المصدر، واللغة، العلاقة، التغطية / النطاق، والحقوق / الصلاحيات. وبرغم أن البيانات الوصفية تُعد أحد المكونات الأساسية لكافة مجموعات البيانات، إلى أنه يجري إهمالها في ممارسات دعم الإتاحة وإعادة الاستخدام والمشاركة للبيانات (Data Curation)، وخصوصاً من الباحثين الذين يجمعون البيانات الأساسية لاستخدامهم الخاص دون مشاركة الآخرين بها.

البيانات (Data)، والمعلومات (Information)، والمعرفة (Knowledge)، والحكمة (Wisdom):

ما يوحد الأنواع المختلفة من البيانات التي تم التطرق إليها في الأجزاء السابقة هو أنها تشكل القاعدة أو الأساس الذي يقوم عليه هرم المعرفة: فالبيانات تسبق المعلومات، والمعلومات تسبق المعرفة، والمعرفة تسبق الفهم والحكمة (Adler 1986; Weinberger 2011). وتتميز كل طبقة من الهرم من خلال عملية التقطير (التقليل، التلخيص والتجريد، المعالجة والتجهيز، التنظيم، التحليل، التفسير، والتطبيق) التي تضيف التنظيم، المعنى، والقيمة من خلال الكشف عن العلاقات والحقائق حول العالم (انظر الشكل ١-١).



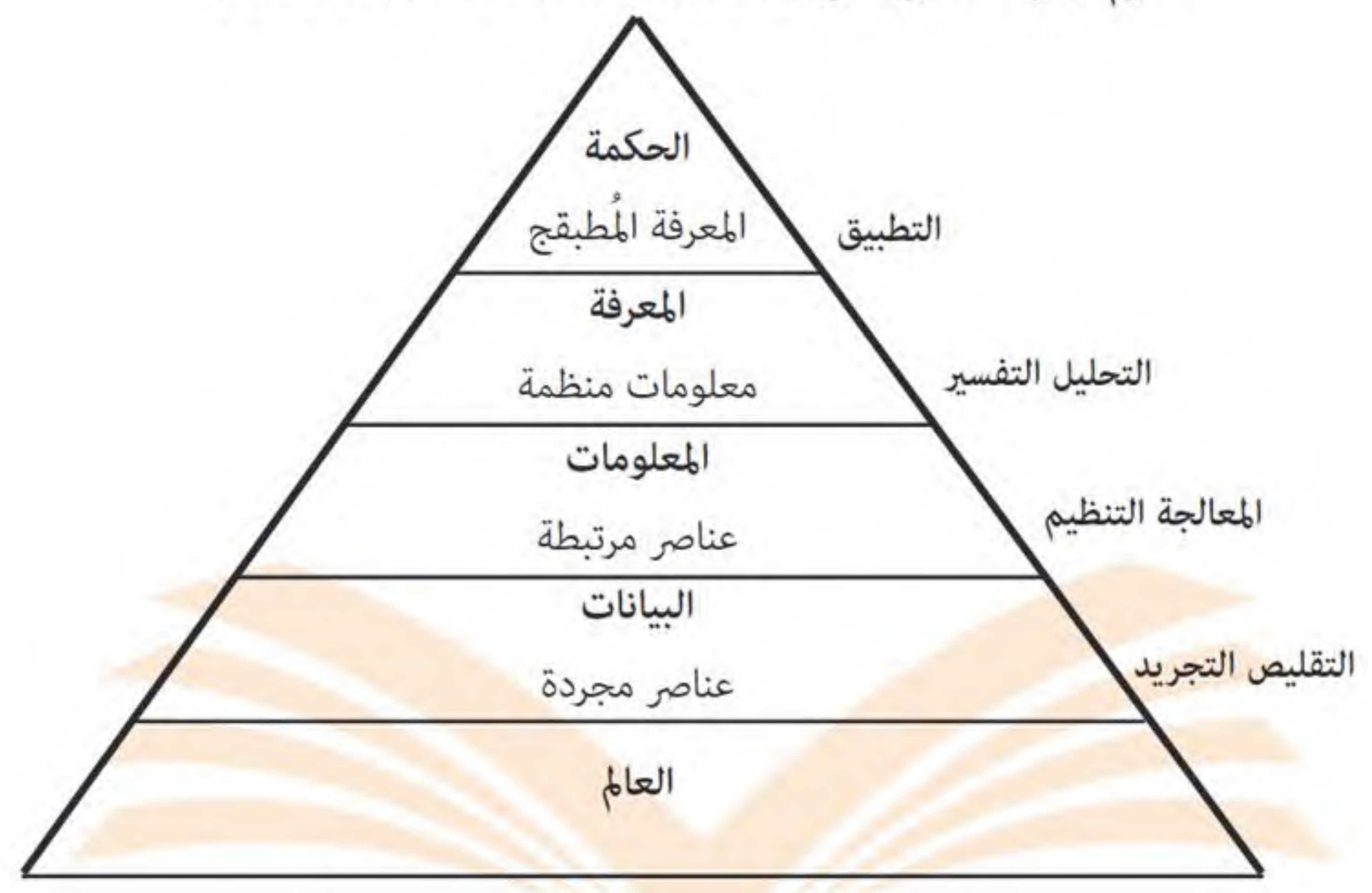
وفي حين أن ترتيب المفاهيم داخل الهرم هو لا جدال فيه على العموم، إلا أن طبيعة المفاهيم والاختلافات بينها كثيراً ما تختلف بين المدارس الفكرية. فالمعلومات، على سبيل المثال، هي المفهوم الذي يفهم على نحو مختلف بين العلماء، فبالنسبة للبعض، المعلومات هي تراكم البيانات المرتبطة بها، وبالنسبة للآخرين هي البيانات إضافة للمعنى، أو الإشارة ذات المعنى في ضوضاء البيانات، أو البناء متعدد الأوجه، أو البيانات العامة حيث تم إعادة صياغة البيانات الأساسية من خلال نموذج تحليلي. فبالنسبة لعالم الفيزياء، البيانات هي مجرد سلسلة من الأصفار والآحاد، أي (Bits) خام؛ وتعد ضوضاء (Noise). ولذلك فإن المعلومات بالنسبة لعالم الفيزياء تتكون عندما يتم تنظيم هذه الأعداد من الصفر والواحد في أنماط متميزة؛ فهذه هي الإشارة المفيدة (von Baeyer 2003). فموجات الأثير وكابلات الاتصالات هي ممتلئة بالمعلومات المتدفقة - كإشارات الراديو والتلفاز والمحادثات الهاتفية وحزم بيانات شبكة الانترنت – التي تعني أنماطا ذات مغزى من البيانات ضمن حزمة أوسع من الضوضاء. وبالنسبة لآخرين، فالمعلومات ذات مفهوم أوسع. فقد حدّد (Floridi 2010: 74)، على سبيل المثال، ثلاثة أنواع من المعلومات :

- الواقعية (Factual): معلومات تعكس الحقيقة (على سبيل المثال، الأنماط، بصمات الأصابع، وحلقات جذوع الأشجار).
- التعليمية (Instrucional): معلومات لأجل الحقيقة (على سبيل المثال، الأوامر، الخوارزميات، والوصفات).
- الدلالية (Semantic): معلومات عن الحقيقة (على سبيل المثال، الجداول الزمنية للقطارات، الخرائط، والسير الذاتية).



شكل (١-١)

هرم المعرفة (منقول من Adler 1986 and McCandless 2010)



فالنوع الأول هو أساساً بيانات ذات معنى، وهو ما يطلق عليه عادة الحقائق. وهي البيانـات التي يتم تنظيمها وتركيبها ضمن نظـام قياس أو مرجع خارجي يوفر بطبيعته أساسـاً لإنشـاء المعنى الأولي الذي يحمل بعض الحقيقة. وتمتـد المعلومات أيضاً إلى ما بعـد البيانات والحقائـق من خلال إضافة قيمة تسـاعد على التفسـير. وعليه، أوضح (Weinberger 2011: 2) أن : "المعلومـات بالنسـبة للبيانات، كالنبيذ بالنسـبة لحقل العنـب: اسـتخراج نواتج التقطير اللذيذة! ". ولذا يمكـن الحصول على هذه القيمة من خـلال فرز، وتصنيف، وربط، وإضافة المحتوى الدلالي من خلال بعض الشـكل النصي أو الصوري والذي يشـير إلى شيء ما و/ أو يرشـد إلى ما ينبغي القيام به (على سبيل المثال، الإشـارة الضوئية التحذيرية على لوحة معلومات مؤشرات عمل السـيارة التي تشـير إلى فـراغ البطاريـة وحاجتها لإعادة الشـحن، Floridi, 2010). وقد بيّن (2002: ملخصاً في Borgman 2007: 40) Case أن الاختلافـات في تعريف المعلومات تتوقف على خمس مسائل:



عــدم اليقــين (Uncertainty)، أو ما إذا كان شيء ما قد يحد من عدم اليقين في تأهيــل البيانــات لتكون معلومــات؛ المادـــة (Physicality)، أو مــا إذا كان شيء ينبغــي أن يؤخذ على شــكل مادي مثل كتــاب، أو كائن، أو موجات صوتية لكلمــة خطيب لتأهل البيانات لتكون معلومــات؛ البنية / العملية (/Structure Process)، أو ما إذا كان يلزم مجموعة من الترتيبات أو العلاقات؛ القصد والعزيمة (Intentionality) أو ما إذا كان شخص ما قد أوعز بأن شيئاً ما تم التواصل معه ليتأهــل ليكون معلومات؛ والحقيقــة (Truth)، أو مــا إذا كان ينبغي أن يكون الشيء صحيحاً للتأهل ليكون معلومات.

وبغــض النظر عن كيفية تصور ذلك، فقد لاحظ (Floridi 2010) أنه بالنظر إلى إضافة المعلومات المعنى إلى البيانات، فهي تكتسب القيمة الربحية كسلعة، غير أنها، كنوع متميز من السلع، تمتلك ثلاثة خصائص رئيسية مميزة (والتي تتميز بها البيانات أيضا):

غير تنافسية (Non-rivalrous): حيث من الممكن لأكثر من مستخدم أن يمتلك نفس المعلومات دون التأثير على مدى اســتفادة المســتخدمين الآخرين منها (على عكس الســلع المادية المستهلكة).

غير قابلة للاســتثناء أو الإقصاء (Non-excludable): حيث يتم تقاســمها ومشــاركتها بســهولة ويلزم بذل جهد كبير في السعي للحدّ من مثل هذه المشاركة (مثل فرض اتفاقيات حقوق الملكية الفكرية أو إضافة حواجز الدفع (Paywalls) قبل الاستخدام).

التكلفة الهامشــية (Zero Marginal Cost): بمجرد إتاحة المعلومات من المرة الأولى، فإن تكلفة الاستنساخ أو إعادة الإنتاج في كثير من الأحيان لا تكاد تذكر.

ومع التشــديد على خصائص كون المعلومات مورد يتميز بعدم المنازعة وعدم الاستثناء، ولأن المعلومات تعد ســلعاً قيمة، تســعى العديد من الكيانات لضبطها والحد من تداولها والســيطرة عليها ومن ثم زيادة قيمتها. ومعظم هــذه القيمة تضاف من خلال العمليات التي تفرض خلال دورة حياة المعلومات (Floridi 2010) والمشتملة على المراحل التالية:



الوجود (Occurrence): الاكتشاف، التصميم، التأليف.

الإرسال (Transmission): التشبيك، التوزيع، النفاذ، الاسترجاع، النقل.

المعالجة والإدارة (Processing and Management): الجمع، التحقق، التعديل، التنظيم، الفرز، التصنيف، التصفية، التحديث، التخزين.

الاستخدام (Usage): الرصد والمراقبة، النمذجة، التحليل، الشرح، التخطيط، التنبؤ، اتخاذ القرارات، التوجيه، التعليم، التعلم.

ومن خلال المعالجة والإدارة والاستخدام يتم تحويل المعلومات إلى معرفة أكثر قيمة.

وكما هو الحال مع جميع المفاهيم في هرم المعرفة، وبالمثل، يختلف فهم مفهوم المعرفة. فبالنسبة للبعض، المعرفة هي الدراية التي تحول المعلومات إلى تعليمات (Weinberger 2011: 3). فعلى سبيل المثال، يمكن ربط المعلومات الدلالية إلى وصفات وتعليمات (أولاً قم بهذا، ثم افعل ذلك..) أو نموذج شرطي لإجراءات استنتاجية (إذا كان الحال هو كذا وكذا فافعل ذاك، وإلا افعل هذا) (Floridi 2010). في هذا التأطير، تكون المعلومات هي البيانات الهيكلية، والمعرفة هي معلومات إجرائية قابلة للتنفيذ (Weinberger 2011). وبعبارة أخرى، "المعرفة هي الوصفة التي تحوّل المعلومات إلى الخبز القابل للأكل، بينما البيانات هي العناصر المكونة للدقيق والخميرة" (Zelany 1987، منقول عن Weinberger 2011). وبالنسبة للآخرين، فإن المعرفة هي أكثر من مجرد مجموعة من التعليمات ؛ ويمكن أن تكون المهارات العملية، وأبعد بكثير عن مجرد وسيلة لمعرفة كيفية إجراء أو تحقيق المهمة، أو النظام الفكري الذي يربط بشكل متسق المعلومات مع بعضها البعض للكشف عن صورة أوسع لظاهرة ما. وينطوي إيجاد المعرفة على تطبيق العمليات المعرفية المعقدة على المعلومات مثل الإدراك، والجمع والتركيب، والاستخلاص، والدمج والتكوين، والاستنتاج والربط. وللمعرفة قيمة أكبر من قيمة المعلومات لأنها توفر أساساً لفهم، وتوضيح، ورسم رؤى حول العالم، والتي من الممكن استخدامها في صياغة السياسات والإجراءات. أما الحكمة، التي تحتل قمة هرم المعرفة، فهي القدرة على تطبيق المعرفة برجاحة وروية.



وبرغم أنه لم تتأصّل بثبات علاقة كل أشكال المعرفة بالبيانات، على سبيل المثال، التخمين والآراء والمعتقدات، إلا أن البيانات تُشكل بوضوح مادة القاعدة الرئيسية لكيفية استشعار الواقـع والعالم من حولنا. حيث تشـكّل البيانــات المدخلات الأساسـية في العمليات مثل الجمع، الفرز، التصنيف، المطابقة، التشـخيص / التنميط، والنمذجة التي تسعى إلى إنشاء المعلومـات والمعرفـة من أجل فهم الظواهـر، والتنبؤ بها، وضبطها والسـيطرة عليها. كما يمكّننـا إنتاج البيانات بمرور الوقـت وفي أماكن مختلفة من تتبع، وتقييم، ومقارنة الظواهر عبر الأزمنة والأماكن والنطاقات. وهكذا، وعلى الرغم من أنه ينظر إلى المعلومات والمعرفة إلى كونها مفاهيم ذات درجة أعلى وقيمة أكبر، تبقى البيانات، مع ذلك، عنصراً أساسـياً مع قيمـة كامنة كبيرة يجرى إدراكها وتحصيلها عنـد تحويلها إلى معلومات ومعارف. فالدافع المنطقـي الرئيسي لحركة البيانات المفتوحة، والتي بحثت في الفصل الثالث، هو القدرة على الوصول إلى القيمة الكامنة في مجموعات بيانات القطاعات الإدارية والعامة.

تأطير البيانات (Framing Data):

حتى الآن في هذا الفصل، فقد بدأت بالفعل صياغة البيانات من الناحية النظرية من حيث مناقشة أنطولوجيا البيانات (ماهية وجودها)، وأشكالها المختلفة، وأين تتموضع داخل هرم المعرفة. كما أن هناك عدداً لا يحصى من الطرق الأخرى التي يمكن بها فهم البيانات والتفكير بها، على سـبيل المثال، من المنظـور التقني الخاص بجودة البيانــات، وصحتها، وموثوقيتها، ونسـبتها، وقابليتها للاسـتخدام، وكيف يمكن معالجتها وتنظيمها، وتقاسـمها، وتحليلها؛ أو من المنظور الأخلاقي المتعلق بالأسـباب التي دفعت إلى إنتاج البيانات والاسـتخدامات التي يتم توظيفها بها؛ أو من المنظور السـياسي أو الاقتصادي والـذي يأخذ بعين الاعتبار التصور المعياري للبيانات ومواضع الطعن والتنازع والتنافس فيها كمنفعة عامة، ورأس مال سياسي، وملكية فكرية، أو سـلعة تجارية، وكيف يتم تنظيمها وتداولها؛ أو من المنظور المكاني و/ أو الزماني المتعلق بكيفية إنتاج البيانات وتطوير استخداماتها ومحاور تركيزها من قبل الأنظمة والمجموعات التقنية، والأخلاقية، والسياسية، والاقتصادية عبر المكان والزمان؛ أو من المنظور الفلسـفي الذي يأخذ بعين الاعتبار المفاهيم المبدئية والمعرفية للبيانات. وسنرجع لاستذكار العديد من القضايا، التي تمّ بحثها في هذا القسم، لاحقاً في جميع أجزاء الكتاب المقبلة.



المنظور الفني (Technically):

في جميع التخصصات، يتم أخذ البيانات بعين الاعتبار من وجهة النظر التقنية المعيارية. والسؤال المطروح على المحك هو إلى أي مدى تنتج طرق التقاط البيانات وقياسها بيانات محددة، ونظيفة، ودقيقة، وكيف يمكن وينبغي معالجة، وتنظيم، ومشاركة، وتحليل هذه البيانات بطرق تحافظ على سلامتها، من أجل ضمان موثوقيتها وصلاحية الاستنتاجات المستخلصة منها. ولذا كان هنالك دائماً شكوك فيما يتعلق بصحة البيانات لأنها بطبيعتها مجرّدة، وعمومية، وتقريبية عند إنتاجها (Goodchild 2009). وحيث ذلك، فقد توجه الكثير من الاهتمام إلى قضايا تتعلق بمفاهيم تمثيل البيانات (Data representativeness)، وعدم اليقين بصحتها (Uncertainty)، ودرجة الموثوقية بها (Reliability)، والأخطاء الممكنة المصاحبة لها (Error)، والتحيز المنتقص لمصداقية دلالاتها (Bias)، والمعايرة (Calibration) في تصميم البحوث وتطبيقاتها، وتسجيل هذه المعلومات كبيانات وصفية.

وبالنظر للبيانات على أنها علامة بديلة عن بعض جوانب ظاهرة ما - كالضوء يمثل نجماً، والخصائص الفيزيائية التي تمثل مبنى، والكلمات التي تمثل أفكار شخص - يتعلق التمثيل (Representativeness) بمدى التقاط البيانات لهذه الظاهرة التي تسعى إلى تمثيلها، وإلى أي مدى تمثل عينة البيانات التي تم إنشاؤها المجتمع الكامل للظاهرة. وفيما يتعلق بالسؤال السابق، فإن السؤال الرئيسي هو إلى أي مدى يمكن أن نكون على ثقة من أن التقنيات العلمية قد التقطت بدقة الظاهرة قيد الدراسة. لقد مثّل هذا التساؤل مشكلة خاصة في مجال العلوم الاجتماعية والإنسانية وأثبت أنه كان من الصعب حلها. فعلى سبيل المثال، لوحظ بشكل كبير أن ما يقول الناس أنهم سيفعلون وما يفعلونه في الغالب ما يكون مختلفاً تماماً عن الواقع، وما يفعله الناس في كثير من الأحيان قد لا يقصدونه. ولذلك برز سؤال حول المدى الذي تمثله بيانات المقابلة الشخصية للسلوك البشري، أو مدى تمثيل الفكر الواعي لهم. وبالمثل، هناك مخاوف بشأن مدى التقاط وتمثيل المؤشرات الرئيسية للتعبير بدرجة كافية عن كيفية أداء مجال ما. على سبيل المثال، إلى أي مدى تُعطي المؤشرات المتعلقة بعدد مرات الاقتباس، ومؤشر إتش (H-index)، وبراءات الاختراع المسجلة دلالة على أداء عالي الجودة من قبل موظفي الجامعة (مع الاحترام لأعضاء هيئة

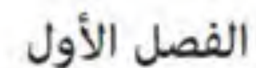


التدريس في مجال العلوم الإنسانية فإن هذه المؤشرات تُعد سيئة للغاية)؟ ولذا كان الحل في محاولة تطوير تصاميم بحثية أكثر تعقيداً لمواجهة أوجه القصور في هذه الأساليب المختلفة، أو تجاهل أوجه القصور المتعلقة بها إلى حد كبير.

وبالنسبة لما يتعلق بمدى تمثيل عينة لمجتمع دراسة ما، فقد نقرر، اعتماداً على خمسين نجمـة، إنتاج بيانـات فلكية هائلـة، ومفصّلة، وعلى مـدى طويل من أجـل فهم أفضل لطبيعتها. ولكن إلى أي مدى يمكن أن نكون على ثقة بأن هذه الخمسين نجمة تمثل مجمل خصائـص المليـارات الأخرى من النجوم الموجودة؟ وحتى في عـصر البيانات الكبيرة، التي تسـعى إلى أن تكون شاملة وليسـت انتقائية في إنتاج البيانات (انظر الفصل الرابع)، فإن البيانات هي بطبيعتها عينة (فليس كل الناس يسـتخدمون وسائل التواصل الاجتماعي، أو يتسوقون باستخدام البطاقات الائتمانية، وبالتأكيد لا يمتلك كثير من الناس في جميع أنحاء العـالم حق الوصول إلى الهواتف وأجهزة الحاسـب)، وهذا يعنـي أن البيانات هي ممثلة لمجموعة من الناس، حتى لو كانت هذه المجموعة كبيرة جداً. ومرة أخرى، كان الحل هو ابتـكار مجموعة من تقنيات أخذ العينات التي تسـعى إلى ضمان التمثيل الدقيق في ظل مختلف الظروف (والتي تعتمد غالباً على عينة عشـوائية)، واسـتحداث أساليب إحصائية تقوم بحسـاب مدى الثقة التي يمكننا أن نكون عليها حيال تمثيل عينة ما لمجتمع الدراسة (Kitchin and Tate 1999).

وتتعلّق الموثوقية بالتكرارية أو الاتسـاق في الحصول على النتيجة نفسـها من توظيف آليـة البحث. وقد وصـف (Golledge and Stimson 1997) ثلاثـة أنواع من الموثوقية: (١) الموثوقيـة التَصَوُّرِيّة (quixotic reliability)، حيـث تؤدي طريقة واحدة للملاحظة والاستكشاف إلى قياس غير متغير باستمرار. (٢) الموثوقية الزمنية المتعاقبة (Diachronic Reliability) وفيهـا يكـون اسـتقرار الملاحظـات عـبر الزمـن. (٣) الموثوقيـة التزامنية (Synchronic Reliability)، وفيها يكون التشابه في الملاحظات خلال الفترة الزمنية نفسها. وتُعد الموثوقية مهمة لأنه من المسلّم به أنه كلما كان القياس أكثر اتساقاً في إنتاج البيانات، أمكن بشكل أكبر وصف هذه البيانات بأنها بيانات موثوقة.





والخطـأ هو الفرق بين القيمة المقاسـة والقيمة الحقيقية، ويمكن أن يشـمل الغيابات البيانات المفقودة، والأخطاء مثل الخطأ في الإدخال / الترميز أو الخطأ في التصنيف أو سـوء تطبيق الأسـلوب التقني، وسوء الفهم. أما التحيز فهو نوع معين من الخطأ، حيث تنحرف البيانات بسـبب نمط ثابت من الخطأ. وعادة ما يحدث التحيز بسبب الطريقة، أو الأداة، أو أسلوب أخذ العينات المستخدم لإنشاء البيانات والتي لها تأثير مفرط على البيانات التي يتم إنتاجها، أو يمكن تقديمها بسـبب موقف إيديولوجـي أو تطلعات الباحث في كثير من الأحيـان بطريقة غير متعمدة (Kitchin 1996). ومن الممكن أن تتسـبب آليات المعالجة مثـل التجميع أو الإجمال في إحداث تحيز من خلال الحدّ من التباين في مجموعة البيانات مما يؤدي إلى أخطاء المغالطة المنطقية للاستدلال البيئي (Ecological Fallacy Errrors)، بمعنـى أنه عـلى افتراض تمثيل القيمة الإجمالية لمجموع الأفراد بدقة على سـبيل المثال إذا كان لدينا شـخصين وزنهما ٥٠ كيلوجراماً وآخرين وزنهما ١٥٠ كيلوجراماً فسيكون متوسط الـوزن الإجمالي هـو١٠٠ كيلوجرام، ولكـن لا أحد في المجموعة يزن هـذا الوزن في الواقع (Kitchin and Fotheringham 1997). ويتعلـق مفهـوم عدم اليقين بمدى ثقة الباحث بخصـوص دقة البيانات وأي تحليل قائم على أساسـها. أما فيما يتعلـق بالبيانات الكمية، فيتصل بدرجة اليقين بالاختبار الإحصائي وذلك بالنظر إلى البيانات المدخلة، وعادة تُحتسب على أساس الاحتمالات ويُعبّر عنها بمستويات الثقة (Goodchild 2009). أما مفهوم عدم اليقين بالنسبة للبيانات النوعية، فيجري في الغالب تقييمه والحكم عليه من قبل خبير بناء على الخبرة السابقة.

وتتمثل دعائم معالجة هذه المخاوف بالإيمان بأن مثل هذه القضايا تنشأ بسبب الضعف البـشري في تصميم البحـث أو أوجه القصور وعـدم كفاءة الأدوات والمعـدات أو الطرق المسـتخدمة والتي من الممكن التغلب عليها من خلال الحلول التقنية. وهذا يعني أنه من الممكن معالجتها عن طريق تحسين نوعية الإجراءات والمعدات المستخدمة، وتطبيق أنظمة التوحيد القياسي (التقييس- Standardization) والتي توفر معايير جودة البيانات المعروفة مثـل تلك التي أقرتها المنظمة الدولية للمعايير آيزو (ISO)، والتي تعمل على إيجاد سـبل للتعويض عن عدم اليقين والخطأ والتحيز في وسائل التحليل المستخدمة.

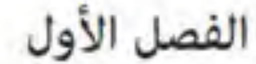


المنظور الأخلاقي (Ethically):

الأخـلاق معنيّـة بالفكر والممارسـة المتعلقة بمفاهيـم القيم مثل العدالة، والمسـاواة، والنزاهـة، والصدق، والاحترام، والحقوق، والواجبات، والرعايـة. ويعمل كل مجتمع وفق مزيج من القيم الأخلاقية الفطرية البديهية والعامية غير الرسـمية، المسـلّم بها، والمواقف الأخلاقية المقننّة جيداً والمنصوص عليها في قواعد، ومبادئ، وسياسـات، وتراخيص، وقوانين تخضـع للإنفـاذ من قبل أجهزة الدولـة والوكالات الأخرى. وكثيراً مـا تكون هذه المواقف الأخلاقية محل نزاع بين مجموعات مختلفة تأخذ وجهات نظر متناقضة عن القيم نفسـها وإلى أي مدى ينبغي إخضاع المواقف الأخلاقية للتشريع القانوني، ومناقشـتها كممارسـة في الفلسفة الأخلاقية. كما يوجد مثل هذا الطعن فيما يتعلق بالبيانات، وخاصة ما يتعلق بما يتم إنشـاؤه من بيانات ووسائل الإنتاج لها، وكيفية مشـاركة البيانات وتقاسمها، وتداولها، وحمايتها، وإلى الحدود التي ينتهي عندها العمل بها.

ففي حـين، تعد بعض البيانات معتدلة نسـبياً، على سـبيل المثال القياسـات المتعلقة بالطقس، فإن بعض البيانات الأخرى تعد حساسة للغاية، كالبيانات المتعلقة بالأفراد والتي يمكن استخدامها لإنتاج صورة مفصّلة عن الحياة المعيشية لهم، وضبط وتنظيم تلك الحياة عـلى إيقـاع معين. وقد يلحـق إنتاج البيانـات ضرراً في بعض الحالات، على سـبيل المثال الأضرار التـي قد تنتـج عن إجراء مقابلات مـع ناجين من جرائم الحرب وما قد تتسـبب فيـه تلـك المقابلات من ضغوطات نفسـية وضيق. وفي هذا المقام، هناك تسـاؤلات حول المـدى الذي من الممكن أن يشـكل فيه إنتاج البيانات، ومختلف أشـكال المراقبة والترصد للبيانات (Dataveillance) أي المراقبة والرصد من خلال معالجة وتحليل سجلات البيانات، وتحليـل البيانات تعديًّا على الخصوصية وحقوق الإنسـان الأخرى، وتسـاؤلات أيضاً حول المدى الذي يمكن أن تستخدم فيه هذه البيانات بفعالية لفرز الأفراد اجتماعياً بمعنى تقديم خدمـات تفضيلية على أسـاس خصائص الأفراد المميزة لهـم (Graham 2005). وتتعاظم هذه المخاوف بالنظر إلى سـهولة جمع ومشاركة وتداول البيانات الرقمية، ونحن نعيش في عصر توليد البيانات ومراقبتها بصورة طاغية وعلى نطاق واسع. ولعله ليس من المستغرب بعد ذلك أن وكالات تمويل البحث العلمي ومؤسسات التعليم العالي الآن تقيّم بشكل دوريّ





الأبعاد الأخلاقية للمشـــاريع البحثية نظراً لآثارها المحتملة على نطاق أوسع، وإضافة لذلك فقد سنّت الدول تشريعات، مثل قوانين حماية البيانات وقوانين الخصوصية، في محاولة لمنع سوء الاستخدام والتجاوزات المسيئة للبيانات، ولقد تمّ مناقشة هذه المسائل وما يتصل بها بشكل مفصل في الفصل العاشر.

المنظور السياسي والاقتصادي (Politically and Economically):

مثـــل الاهتمام بالمنظور الأخلاقي للبيانات البداية للكشـــف عن الطرق التي تعمل على تأطـــير البيانات من خلال الاهتمامات السياســـية والاقتصادية على نطاق أوســـع. وتتعلق هذه الاهتمامات بماهية البيانات التي يتم إنشـــاؤها، وكيفية معالجتها، وتحليلها، وتوظيفها بالنسبة إلى: كيفية تصوّر البيانات وتوصيفها معيارياً ضمن المجموعات السكانية وتوظيفها من قبل الدول، ومفاهيم الكيفية التي ينبغي بها تنظيم البيانات وتقنينها؛ والمناقشات التي تطرح ضمن مجموعات الاختصاص التي تؤيد أو تعارض إنتاج البيانات وتطبيقاتها؛ واتخاذ القرار حول التمويل والاســـتثمار في البيانات؛ وانتشـــار الرأســـمالية والطرق التي تســـتخدم البيانات لإدارة المساعي وزيادة القيمة والأرباح؛ والتداول للبيانات كسلعة مع ظهور سوق للبيانات تُقدر قيمته بمليارات الدولارات والتي يشـــارك فيها مجموعة متنوعة من اللاعبين كالمنتجين، المجمّعين، البائعين، المحللين، والمســـتهلكين، انظر الفصل الثاني. فمنتجو البيانات ينبغـــي عليهم الأخذ بعـــين الاعتبار الرأي العام والســـياسي، والاعتبـــارات الأخلاقية والبيئة التنظيمية والتمويل المتاح وسلامة استثماراتهم إزاء الموارد. كما ينبغي على أولئك القائمين على الساحتين التشريعية والتمويل المالي أن يكون لديهم سعة الأفق واتخاذ القرارات حول كيفية تشـــكيل المشهد الذي يعمل به منتجو ومستخدمو البيانات، وكذلك النظر في أنظمة العمـــل المتعلقة بالبيانـــات الخاصة بهم وما يكشـــفونه عن جداول أعمالهـــم وأولوياتهم وأساليب الإدارة والحوكمة الخاصة بهم (Lauriault 2012).

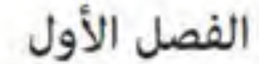
وفي كلتا الحالتين، هناك مجموعة متنوعة من العلاقات السياسية والاقتصادية العقلانية في المشـــهد، مع أصوات متنافســـة تســـعى للتأثير على الرأي ومشـــهد البيانات على النطاق الأوسع. فعلى سبيل المثال، فإن حركة البيانات المفتوحة تمثّل البيانات كسلعة عامة تتشكّل مـــن العموم وتكـــون في متناول الجميع بحرية مطلقة باســـتثناء البيانات الحساســـة أمنياً،



والشخصية الخاصة أو يمكن الوصول إليها من خلال اتفاقيات الاستخدام العادل. وفي المقابل، يعرض قطاع الأعمال البيانات كسلعة تجارية قيّمة تحتاج من جهة إلى الحماية من خلال أنظمة الملكية الفكرية (حقوق الطبع والنشر، براءات الاختراع، وحقوق الملكية) ولكن، من جهة أخرى لا ينبغي ربطها بشدة بالمحاذير الأخلاقية التي تمنع من استغلالها لتحقيق المكاسب المالية. أما بالنسبة للمجتمعات والدول، فالبيانات هي الوسائل التي يمكن من خلالها تقنين، وممارسة، ومنازعة الأجندات السياسية والأعمال عبر تمكين بناء الإثباتات - أي الروايات المسندة والدعوات المضادة التي لها قيمة بلاغية أكبر من مجرد التندر أو دغدغة المشاعر (Wilson 2011; Garvey 2013). وبعبارة أخرى، تعد البيانات بالنسبة لعالم الفلسفة الفرنسي Foucault's 1981) شكلاً من أشكال السلطة / المعرفة؛ أي وسيلة يمكن من خلالها ضبط وتنظيم الشعوب، والظواهر، والأقاليم (Lauriault 2012). وتتقاطع هذه الاهتمامات والمصالح البديلة في كثير من الأحيان بطرق متناقضة، من حيث أنه قد يكون لها أجندات مختلفة تماماً، على سبيل المثال دعم الأعمال التجارية الكبيرة لحركة البيانات المفتوحة فيما يتعلق بالبيانات العامة (انظر الفصل الثالث). وبعبارة أخرى، تظهر البيانات جلياً وتتموضع ضمن الاقتصادات السياسية المعقدة، وفي الوقت نفسه فهي تستخدم لتشكيل مثل هذه القضايا والنظم والمجموعات.

وعلاوة على ذلك، تشكل البيانات مورداً اقتصادياً، وهي عنصر أساسي في المرحلة المقبلة من الاقتصاد القائم على المعرفة، وتعمل على إعادة تشكيل نمط الإنتاج القائم على البيانات (انظر الفصل السابع). ومنذ أواخر الثمانينيات (١٩٨٠s)، يرى بعض العلماء مثل Castells (1988, 1996) أن آخر دورة للرأسمالية يدعمها إنتاج المعرفة التي تنشئ منتجات وأشكال عمالة جديدة، وتسهّل إعادة الهيكلة الاقتصادية، وتعزز الإنتاجية والقدرة التنافسية، والكفاءة، والاستدامة، وتراكم رأس المال. وتُعد البيانات الكبيرة، على وجه الخصوص، هي أحدث التطورات المتعلقة بتعميق وتعزيز هذه الدورة، وتوفير ثورة من الأدلة التي يتم استخدامها من قبل الشركات من جهة، لرصد وتقييم أداء الشركة اللحظي، والحدّ من الهدر والاحتيال، وتحسين إستراتيجية الشركات، والتخطيط وصنع القرار، ومن جهة أخرى، لتصميم سلع جديدة، وتحديد واستهداف أسواق جديدة، وتطبيق التسعير المتغير، وإدراك الإمكانات غير المستغلة، واكتساب ميزة تنافسية (Manyika et al. 2011; Zikopoulos et al. 2012). وفي سبيل





القيام بذلك، يمكّن إنتاج وتحليل البيانات من تشغيل الشركات بشكل أكثر ذكاء فيما يتعلق بكيفية تنظيمها وتشغيلها، وتعزيز المرونة والابتكار، والحد من المخاطر والتكاليف والخسائر التشغيلية، وتحسين تجربة العملاء، وتعظيم العائد من الاستثمار والأرباح. وكقائد لعملية تراكم رأس المال، ستسهم البيانات الكبيرة في إيجاد أعمال (وظائف) جديدة وفي الجولة المقبلة من التنمية غير المتوازنة، ومن هذا المنظور يمكن فهم البيانات على أنها وكيل لمصالح رأس المال واهتماماته.

المنظور الزماني والمكاني (Temporally and Spatially):

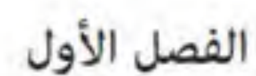
تمتلك البيانات أبعاداً زمانية ومكانية على حدّ سواء. حيث تختلف ماهية البيانات التي يتمّ إنتاجها والطرق التي يتمّ بها معالجتها، وتحليلها، وتخزينها، أو التخلص منها باختلاف الزمان والمكان؛ أي أن للبيانات ومجاميعها المحيطة بها دلالات تاريخية وجغرافية. كما تتغير كيفية معالجة البيانات وتحليلها مع مرور الوقت، متأثرة بالتغيرات والتحسينات التنظيمية في التعداد والإدارة، والقوانين الجديدة المتعلقة بتداول البيانات وحمايتها، والتقنيات الجديدة، والطرق الجديدة في فرز البيانات وتحليلها، والاختلافات الإحصائية الجغرافية مثل حدود المقاطعة المحلية أو الوطنية الجديدة، والأساليب الإحصائية الجديدة. وعلاوة على ذلك، يمكن لمجاميع البيانات التي تستخدم في ولاية معينة أن تكون مختلفة تماماً عن تلك التي تستخدم في أخرى، وحتى داخل حدود الولاية الواحدة، فقد تختلف الطريقة التي يقوم فيها كيان ما بإنتاج البيانات وإدارتها بسبب التقلبات المؤسسية أو الشخصية.

وبالنظر في التعدادات السكانية، يتكوّن التعداد من مسح شامل لمنطقة وساكنيها، ويتم إجراء المسح في العادة كل عشر سنوات، والهدف من ذلك هو جمع معلومات أساسية حول المقيمين في منطقة ما وخصائصهم على سبيل المثال، العمر، الجنس، الحالة الاجتماعية، التركيبة الأسرية، الدين، العرق، الطبقة الاجتماعية، وما إلى ذلك وجوانب عيشهم فيها (عملهم، ومكان الإقامة، وما إلى ذلك). ويتطلب تمكين قياس التغيير في التعدادات السكانية الاستمرارية فيما يتعلق بالأسئلة المطروحة فيها والكيفية التي تدار بها هذه التعدادات. وفي ذات الوقت، يتطلب التقاط البيانات الجديدة موضع الاهتمام



والتـــي تعكس تغيرات أوســع في المجتمع، إجراء تغييرات على آليــة هذه التعدادات، مثل إضافة أسئلة جديدة أو التعديل على الأسئلة الأساسية (انظر الشكل ٢-١: ولاحظ كيف أنه حتى عندما تمّ الحفاظ على الأسئلة عبر التعدادات السكانية المتعاقبة، فإن كيفية صياغتها كانــت غالبــاً ما تكون مختلفــة تماماً في كل مــرة). وعلاوة على ذلك، فــإن الكيفية التي يدار بها التعداد الســكاني كانت تتشــكّل في التعدادات المتعاقبة من قبل قوى مؤسســية، وسياســية، واقتصادية، وتطورات تقنيــة جديدة : انظــر Linehan (1991) للتعرف على تاريخ التعداد السكاني الايرلندي من ١٨٢١ – ١٩٩١، و2012) Lauriault) لتحليل التعداد الكنــدي من ١٨٧١ - ٢٠١١. عــلاوة على ذلك، يتعرض التعــداد للطعن والتفاوض حيث تتنافس مصالح خاصة لضمّ، أو تغيير، أو إزالة الأســئلة منها. وفي بعض الحالات، يمكن أن تكون التغييرات المطلوبة جذرية بشــكل متطرف، مثل القرار الألماني بوقف إجراء التعداد السكاني في الثمانينات (انظر Hannah 2011). ونتيجة لذلك، يرزح التعداد الوطني دائماً تحت وطأة التوتر الناتج عن الاســتمرارية والتغيير، ومع ذلك، يتطور مع مرور الوقت وفي مناطــق جغرافية مختلفة. وحتــى الآن، ومع ذلك، فقد كان هناك عدد قليل من مجاميع البيانــات المميزة تاريخياً وجغرافياً (;Alder 2002; Desrosières 1988; Hannah 2011 .(Hewitt 2010; Lauriault 2012; Poovey 1998; Porter 1995





شكل (٢-١)

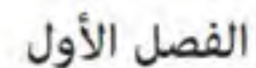
الأسئلة المتعلقة بالأفراد خلال التعداد السكاني الايرلندي للسكان ما بين ١٨٤١ - ١٩٩١

س	السنة	١٨٤١	١٨٥١	١٨٦١	١٨٧١	١٨٨١	١٨٩١	١٩٠١	١٩١١	١٩٢٦	١٩٣٦	١٩٤٦	١٩٥١	١٩٥٦	١٩٦١	١٩٦٦	١٩٧١	١٩٧٩	١٩٨١	١٩٨٦	١٩٩١
ش	الشهر	٦	٣	٤	٤	٤	٤	-	٤	٤	٤	٥	٤	٤	٤	٤	٤	٤	٤	٤	٤
ي	اليوم	٦	٣٠	٧	٢	٣	٥	-	٢	١٨	١٨	١٢	٨	٨	٩	١٧	١٨	١	٥	١٣	٢١
ع	سعة النماذج/ أعداد الأشخاص	٢٠	١٥	١٥	١٢	١٥	١٥	-	١٥	١٠	١٠	٩	١٠	١١	١٠	١٠	١٠	١٠	٦	٨	٨
١	الاسم الأول وأسم العائلة	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
٢	الجنس	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
٣	العلاقة بالنسبة لرب الأسرة	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X	X	X	X	X
٤	العمر وتاريخ الميلاد	A	A	A	A	A	A	A	A	B	B	C	C	C	C	C	C	C	C	C	C
٥	الحالة الاجتماعية	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	Y	Y	YS	YS
٦	مكان الميلاد	X	X	X	X	X	X	X	X	X	X	X	-	-	X	-	X	-	X	X	X
٧	أعضاء الأسرة الغائبين	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
٨	حالات الوفاة في الأسرة منذ آخر تعداد	X	X	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
٩	مستوى الأمية (معرفة القراءة والكتابة)	Y	Y	Y	Y	Y	Y	Y	Y	-	-	-	-	-	-	-	-	-	-	-	-
١٠	الوظيفة	X	X	X	X	X	X	X	X	X	X	X	X	-	X	X	X	-	X	X	X
١١	تاريخ الزواج	D	D	D	D	-	-	-	dw	dw	-	dw	-	-	dw	-	dw	-	dw	-	
١٢	القدرة على تحدث اللغة الايرلندية	-	*Y	*Y	*Y	Y	Y	Y	Y	Y	Y	Y	-	-	Y	-	Y	-	Y	Y	Y
١٣	الإعاقة	-	D	D	D	DL	DL	DL	DL	-	-	-	-	-	-	-	-	-	-	-	-
١٤	الديانة	-	-	X	X	X	X	X	X	X	X	X	-	-	X	-	X	-	X	-	X
١٥	عدد الأبناء من آخر زوجة	-	-	-	-	-	-	-	X	X	-	X	-	-	X	-	X	-	X	-	X
١٦	عدد الأبناء من آخر زوجة على قيد الحياة	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-	-	-	-	-	-
١٧	التيتم (للأطفال)	-	-	-	-	-	-	-	-	X	X	X	-	-	-	-	-	-	-	-	-
١٨	عدد التابعين	-	-	-	-	-	-	-	-	X	-	X	-	-	-	-	-	-	-	-	-
١٩	مجال العمل (صاحب عمل أم موظف)	-	-	-	-	-	-	-	-	X	X	X	X	-	X	X	X	-	X	X	X
٢٠	المنطقة	-	-	-	-	-	-	-	-	H	H	H	H	-	H	I	H	-	I	-	I
٢١	فترة البطالة للعام السابق	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-	-	-
٢٢	مدة الإقامة للمهاجرين	-	-	-	-	-	-	-	-	-	-	X	-	-	X	-	-	-	-	-	-
٢٣	عنوان المنزل للزوار	-	-	-	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-	-
٢٤	الوظيفة	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X	-	Y	Y	Y
٢٥	الوظيفة الفرعية	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-
٢٦	سن التوقف عنده عن الدراسة النظامية	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	-	X	X	X
٢٧	أنواع المدارس	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	IY	-	IY	-	-
٢٨	المؤهلات العلمية أو التكنولوجية	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	X	X
٢٩	الإقامة المعتادة الآن	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	X	X
٣٠	الإقامة المعتادة قبل عام واحد	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	X	X
٣١	وسائل التنقل للعمل أو المدرسة، الخ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	Y	Y	Y
٣٢	المسافة إلى العمل، المدرسة، الخ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	-	X	X	X
٣٣	الإقامة في العام الماضي	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	-	-	-
٣٤	عنوان مكان العمل، المدرسة، الخ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X	X
٣٥	الوضع الحالي	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y	Y	Y
٣٦	الإقامة في الخارج	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	X	X
٣٧	المستوى تعليمي	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y
٣٨	الأملاك	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Y

المصدر : منقول بتصرف عن Linehan 1991.

المنظور الفلسفي (Philosophically):

بالنسـبة للبعض، تعد البيانات على المسـتوى الوجودي (Ontological level) حميدة (غـير خطـرة). فالبيانات هي مجرد بيانات، وعناصر أساسـية يتم اسـتخراجها من العالم بطـرق محايدة وموضوعية تخضع لقيـود تقنية. فهي "لا تمتلـك أي معان ضمنية كامنة فيها، ولا تقدّم بالضرورة أي تفسـيرات أو آراء، ولا تحمل خصائص دلالية قد تكشـف عن أهميتهـا وصلاتهـا " (Pérez- Montoro and Díaz Nafría 2010)، فهي تخمينية قابلة للتحليل. ومن هذا المنظور، فإن جهاز الاستشعار (sensor) على سبيل المثال ليس لديه أي سياسـة أو جدول أعمال. فهو ببسـاطة يقيس الضوء أو الحرارة أو الرطوبة، ...الخ – حيث يقوم بإنتاج القراءات التي تعكس حقيقة الواقع عن العالم المحيط ما لم يشـوبه خلل فني. وبعبـارة أخـرى، تُنتج أجهزة الاستشـعار نظرة موضوعية واقعية عن العالم وتكشـف عن الأشـياء كما هي في الواقع، حيث إن قياس واقعية الشيء هو مستقل عن عملية القياس له (Desrosières 1998). أمـا في إطار عمليات القياس التي يلعب فيها العنصر البشري دوراً مركزيـاً - في التجارب المخبرية، أو إجراء مسـح ميداني أو مقابلة شـخصية – فإن عمليات القياس تطبّق شـكلاً من أشـكال الموضوعية الميكانيكية التي تلتزم بقواعد محدّدة، وطرق منهجيـة دقيقة لإنتـاج بيانات منفصلة، غير متصلة، ونزيهة على نحو شـفاف، وخالية من تحيز الباحث وهواه وتفضيلاته، ومن ثم فهي مسـتقلة عـن العادات، والتقاليد، والثقافة، والمعرفة، والسـياق (Porter 1995). وعلى هذا النحو، فإن العلم عند ممارسـته بشـكل صحيـح لا ينبغي أن يتلبس بأي دوافع سياسـية أو أجندات خفيـة بحيث يمكن بعد ذلك أن تؤخذ البيانات على ظاهرها وعلاتّها. وفي الواقع، تقترح المصطلحات المسـتخدمة عادة لتفصيـل كيفية التعامـل مع البيانات عمليات تقنية حميدة مثـل: «الجمع»، «الإدخال»، «الترجمة»، «التخزين»، «المعالجة»، و«التنقيب» (Gitelman and Jackson 2013). ولذا فإن اسـتخدامات البيانات هي التي تخضع للإملاءات السياسـية. وبعبارة أخرى، إن الناس هم من يفسـدون البيانـات ويعملون على تجييرها لمصالحهم الخاصـة، وليس للعلم بحدّ ذاته.

وبالنسبة لآخرين، فإن هذا الرأي لا يمكن تبريره والدفاع عنه، إذ تتحدّد طبيعة البيانات بفعالية من خلال الكيفية التي يمكننا بها تصوّر البيانات، وقياسها، والتعامل معها. فبالنسبة لهم، البيانات لم تكن موجودة قبل إنتاجها؛ إلا أنها لا تنشأ من العدم. ويتمّ إنتاج البيانات مـن خلال القيـاس، والتجريد والاقتبـاس، وتعميم التقنيات التي تـمّ وضعها ضمن تصوّر لتنفيذ مهمة ما وتسـجّل في الأشـكال والمقاييس التي تتفق مع المعايير التي اخترعها الناس على سـبيل المثال، النظام المتري. فالبيانات هي وحدات معرفية، جُعلت على نموذج تمثيلي لتمكـين وإتاحة العمـل المعرفي، وقياس وتسـجيل البيانات عن نفـس الظاهرة من خلال العديد من الطرق والأساليب، التي يقدّم كل منها مجموعة مختلفة من البيانات التي يمكن تحليلها وتفسيرها باستخدام وسائل مختلفة (Poovey 1998). فالكيفيات التي يتمّ إنشاء البيانات بها لا يمكن حصرها نظراً لتعدد التصاميم والنقاشـات والتعديلات التي تطرأ على البروتوكولات، والعمليـات التنظيمية، ومعايير القياس، والفئات، والمعايير الخاصة بالبيانات في ظـل الفـوضى المصاحبة لعملية توليد هـذه البيانات وإنتاجها. فلـو أخذنا حالة قياس التعداد السكاني لبلد ما: فإن هنالك العديد من القرارات التي يجب أن تُتخذ مثل تلك التي تحدّد من يضاف للتعداد ومن يسـتثنى منه على سـبيل المثال، أن يشمل الزوار، والأجانب المقيمين بصورة مشروعة أو غير مشروعة، وأولئك الذين يتجنبون المشـاركة عن عمد أو لا، وما إلى ذلك وأين يجب أن يتمّ تعدادهم على سـبيل المثـال، تعدادهم تبعاً لإقامتهم ليلة إجـراء التعـداد أو مكان إقامتهم الدائمة في العادة؛ فجميع أنـواع القواعد والإجراءات يتم وضعهـا قبل التعداد، ومع ذلك يظل هناك أمور لم يتخـذ قرارات بخصوصها واختلاف عبر العدّادين في طريقة التطبيق (Porter 1995).

ونتيجة لذلك، فإن الكيفية المتبعة لتعريف البيانات وتحديدها وجودياً ليسـت عملية فنيـة محايدة، ولكـن معايرة سياسـية وأخلاقية يجري التنـازع عليها، ولهـا نتائج قابلة للتحليـل، والتفسـير، والعمل في وقـت لاحـق (Bowker and Star 1999). ومع ذلك، لمـرة أخرى، تعمل تبويبات البيانات على اسـتبعاد التباينات والتقلبـات وتحديد ما يتاح ويحجـب داخل مجموعة البيانات، وعلى الرغم من ذلك، تكون هذه العملية من التقارب والاسـتقرار معرضة على الدوام للممانعة والرفض وإعادة الصياغة بسـبب تعدد الجهات الفاعلـة، والموضوعات والعمليات موضع المناقشـة والرد، والتناقـض الحاصل في البيانات



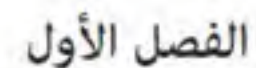
والذي لا يمكن تطبيقه بســهولة وبما يتناســب مع نظام ما (Bowker and Star 1999). وعــلاوة على ذلك، فإنه بمجــرد إنتاج البيانات يكون من الممكن فرز، وتقســيم، وتكعيب هــذه البيانــات بطرق متعددة في عدد من الفئات. وبعبارة أخرى، فإن البيانات ليســت مســتقلة عن نظام الفكر والأدوات التي تدعم إنتاجها (Bowker and Star 1999)، فهذه النظم الفكرية مُطعّمة بافتراضات فلســفية ومعتقدات تُمارس بشــكل مختلف. فالواقع، وكما لاحظ Borgman (2007: 38) أنه في الوقت الذي يسعى فيه العلم إلى تصوير نفسه بالعالمية والعموم، تبدو ممارســات العلوم محلية ومتباينة على نطاق واســع، ومن ثم فإن البيانات لا تمثّل ببســاطة واقع وحقيقة العالم؛ فالبيانــات هي وحدات بنيوية حول العالم (Desrosières 1998; Poovey 1998). مــن هــذا المنظور «يتمّ إنتاج المعرفة العلمية - بدلاً من براءة اكتشافها» '(Gitelman and Jackson 2013: 4). كما في التالي:

> إذا كان إنتــاج البيانــات عن طريق أداة مثل شــبكة استشــعار، فإن تفســير هذه البيانات يتطلب فهم الأداة - على ســبيل المثال، ما الذي تستكشــفه أجهزة الاستشــعار؟ وتحت أي ظروف؟ وكم عدد تكــرار مرات الملاحظة؟ ومع أي نوع من المعايرة؟ (Borgman 2007: 183).

ومــع ذلك نجــد أن العلم يحاول في كثــير من الأحيان جعل البيانات حرة ومســتقلة عن مثل هذا الفهم الســياقي، وتنظيم ومشــاركة البيانات من خلال قواعد البيانات التي يجري التحســين فيها والتخفيف من حدّة الفوضى عند إنشــائها وحماية المستخدمين من معرفــة كيفية إنتاج البيانات وتنظيمها (Gitelman and Jackson 2013). ولذلك، يُبيّن Ribes and Jackson (2013: 165) أن الاســتخدامات العلمية للبيانات والقائمة على أنها محايدة وموضوعية هي مجرد تخيلات لا حقيقة لها؛ أنها «تفترض وتصور العالم بناء على البيانــات المتداولة دون النظر إلى مصدرها، وحالتهــا، وجوهرها، وتاريخها، ومن ثم تبرز هــذه البيانات على أنها حرة للتنقل عبر العالم كعملة غــير متمايزة وعالمية». في المقابل، هما يؤكدان أن «البيانات هي أكثر ارتباطاً من ذلك».

ونتيجة لذلك، فالبيانات غير مســبقة التحليل، وليست عناصر موضوعية مستقلة. كما بيّنها Gitelman and Jackson (2013: 2)، متبعين في ذلك بوكر Bowker (2005) عندما





ذكـروا أن "البيانـات الخام هي مصطلح متناقـض؛ «فالبيانات هي على الـدوام "مطهوة" سـلفاً، ولم تكن يومـاً "خاماً" أبداً. فهناك حاجة لتصـوّر البيانات كما هي على الحقيقة في الوجـود والعمل بموجـب ذلـك" (Gitelman and Jackson 2013: 3). فالبيانات – على حـدّ سـواء - هي اجتماعيـة من باب أنها تقع ضمن السـياق، ومادية من بـاب أن لديها شـكلاً (كأرقـام ثنائية، أو رموز، أو أعـداد.. وما إلى ذلك)، ويجـري تخزينها على الورق، والأشرطـة الممغنطـة، والأقراص الصلبة، ومـا إلى ذلـك (Wilson 2011; Gitelman and Jackson 2013). كلا المنظوريـن يعمل على تشـكيل القواعد التأسيسـية للبيانات، فعلى سـبيل المثال، تعتمد البيانات الكبيرة على النقاشـات الأخلاقية، والاقتصادية، والسياسية في المجالات المختلفة التي تدعم البدء في اعتمادها وتنفيذها (انظر الفصل السـابع)، وتعتمد على قواعد البيانات، والحواسـيب، والخوادم، وشبكات الاتصال التي تمكّن بدورها عمليات إنتاج البيانات، ومعالجتها، ومشـاركتها، وتحليلها، تخزينها (انظر الفصل الخامس). وتسهّل هذه الوسـائط فسـاد البيانات (تعفّنها)، وسوء تموضعها، أو نسـيانها، أو حذفها، أو تآكلها من خلال تلف الوحدة التخزينية للأرقام الثنائية – بت (أي تعطّل وسـائط التخزين نتيجة تحلّـل الأشرطة الممغنطـة أو تضرر الأقراص الصلبـة) (Boellstorff 2013). وفي الواقع لم تكن البيانات قطّ مطهوّة فقط، ولكنها أيضاً منفتحة لتقبّل أساليب «غير مخطط لها، وغير متوقعة، وعرضية»، و«متحولة بأشـكال شبة إنسـانية معقّدة، وطرق زمانية ناشئة لا تتبع على الدوام وصفة حسابية محددة مسبقاً» (Boellstorff 2013).

ونظراً للطبيعة الاجتماعية والمادية للبيانات، فإننا ننشـط في إعادة صياغة وتشكيل «عوالمنـا المادية، والتقنية، والجغرافية، والتنظيميـة، والاجتماعية إلى نوع من البيئات التـي يمكن للبيانات أن تزدهر فيهـا... والوصول إلى العلاقـة التكافلية مع البيانات» (Ribes and Jackson 2013: 152). ومن ثم، وبينما يُنظر إلى البيانات بوصفها شيئاً موجوداً – شـيئاً حول الواقـع (Shah 2013، التأكيد الأصلي)، فإن الفهم الأكثر إنتاجية للبيانـات هو اعتبارهـا أحد مكوّنات الواقع وأحد مُنتجـي الواقع في ذات الوقت. إن البيانات ليسـت محض فكرة تمثيليـة مجردة، بل هي تأسيسـية، وإنتاجها، وتحليلها، وتفسـيرها له نتائج وعواقب تتبعها. وكما بيّن Gitelman and Jackson (2013: 2): «إذا كانـت البيانـات تخضع لنا بطريقة أو بأخرى، فإننا نحـن أيضاً نخضع للبيانات».



حيــث يتمّ التقاط البيانات من العــالم، ولكن في المقابل تعمل في العالم. إن البيانات لم تكــن يوماً، ولا يمكن أن تكون، حميــدة (Shah 2013). بدلاً من ذلك، من الضروري فهم «البيانات على أنها مؤطرة متأثرة بغيرها ومؤطرة مؤثرة على غيرها» (Gitelman and Jackson 2013: 5). وبعبــارة أخــرى، هنــاك حاجة إلى مــا هو أكثر من مجرد الاعتراف العام من قطاعي العلوم والأعمال لأهمية التصوّر المفاهيمي للبيانات.

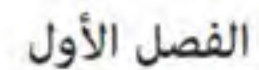
التفكير النقدي حول قواعد البيانات والبنى التحتية للبيانات:

في العادة، يتمّ تجميع البيانات في مجاميع (Datasets) في محاولة لفهم البيانات، والتي غالبــاً ما يتم تنظيمهــا وتخزينها في قواعد بيانات (وهي وســيلة لتنظيم وتخزين البيانات التــي يمكن الاســتعلام عنها بطرق متعــددة) وبنى تحتية للبيانات مثــل نظم المحفوظات الأرشــيفية (Archives) وحاويات البيانات (Repositories) (انظر الفصول الثاني والثالث والخامس). وكما كان الحال مع التصور المفاهيمي للبيانات بشــكل عام، فمن المهم كذلك التفكير بشــكل نقدي حول طبيعة قواعد البيانات والبنى التحتيــة للبيانات، وإنتاجها من الناحيتــين الاجتماعية والتقنيــة (Socio-Technical)، وكيف أنهــا تعكس العقلانية حول العالم في الوقت نفســه لأنها تتكاثر وتعزز هذه العقلانية. وكان هذا التفكير النقدي غائباً إلى حد كبير فيما يتعلق بالبيانات الكبيرة، والبيانات المفتوحة، والبيانات الصغيرة الموسّــعة، مع التركيز حتى الآن عليه ليكون ذا طبيعة تقنية وآلية بشكل أكبر.

وتمتدّ الفرضية المطروحة التي تمّ تبنيها وتطويرها في جميع أنحاء هذا الكتاب من الحجة المبنية في القســم الأخير، وهي الافتراض بأن قواعد البيانات والبنى التحتية للبيانات ليســت ببساطة محايدة، أو وسائل تقنية لتجميع وتبادل البيانات؛ وليست بالتأكيد منتجات لتخزين البيانــات الملتقطة عن العالم، ولكنها حزم مــن العمليات الطارئة والمترابطة التي تقوم بتأدية الأعــمال في العالم (Star and Ruhleder 1996; Kitchin and Dodge 2011). إنها أنظمة اجتماعيــة تقنية معقدة كامنة كجزء لا يتجزأ ضمن المشــهد المؤسســاتي الأكبر من الباحثين، والمؤسسات، والشركات، التي تشكل أدوات أساسية في إنتاج المعرفة والحكم ورأس المال.

ويتــمّ تصميم وبناء قواعــد البيانات على تقبّل أنواع معينة مــن البيانات وتمكين أنواع معينــة من التحليل، ولذا فهنالك آثار عميقة مترتبة على الكيفية التي يتم بها هيكلتها على





الاستفسارات والتحليلات الممكن أداؤها؛ كما أن شروط صياغتها من شأنها أن تحدد العمل الـذي يمكن القيام به عليهـا ومن خلالها (Ruppert 2012). فعلى سـبيل المثال، يختلف تخزين البيانات في قواعد البيانات اختلافاً تاماً عن الشكل السردي من حيث ما تمّ تخزينه، وكيف يتم تخزينه، وكيف من الممكن اسـتدعاؤها وإعادة اسـتخدامها (Bowker 2005). وينتـج عن قواعد البيانات إهمال وتجاهل لبعـض البيانات، إضافة إلى الثغرات الكامنة في البيانات، وذلك كلما تم الكشـف عن علاقات بين البيانات وتوفير الإجابة عن بعض الأسـئلة التي يتعين الإجابة عنها؛ فهي تقيّد وتتيح من خلال وجودها (الأنطولوجيا)، وتُنتج العديد من العلاقـات في ذات الوقت الذي يتم فيه إخفاء علاقات أخرى (Ruppert 2012; Vis 2013). واسـتناداً إلى Derrida، فقد جـادل Bowker (2005: 12) بأن قواعد البيانات والمحفوظات الأرشـيفية هي المتحكم (jussive): فهي "تقول لنا ما نستطيع وما لا نستطيع أن نقوله" من خلال تحديد ما يمكن تذكره وما يجري تجاهله ونسـيانه. فهذا التذكر والتناسي يتم تحديده في الأصل من خلال أنظمة وممارسات مبنية على قوانين سياسية وفلسفية.

ومن هنا فإن التبويبات داخل قواعد البيانات غير ثابتة وغير طبيعية، ولكن يتمّ إنشاؤها مـن قبل جهـات فاعلة ذات أهداف محددة تعمل ضمن مجتمعات الممارسـة، أسـاليب الحوكمـة، والقيود التقنية الفنية. ومن ثم، فإن قواعد البيانات هي تعابير عن السـلطة / المعرفـة وهي تقوم بإقرار تشريعها وإعادة إنتاج هـذه العلاقات (Ruppert 2012)، على سـبيل المثال، تحديد معدل التأمين لشـخص ما أو ما إذا كان بإمكانه السـفر بين البلدان. وعـلاوة على ذلك، فإن قواعد البيانـات هي كيانات ديناميكية متغيرة تقوم بإجراء «عملية مسـتمرة من التمييز والتفريق» (Ruppert 2012: 129) من خلال التفاعل مع مجاميعها المرتبطة بها (المنشئون، المستخدمون، البرمجيات، الأجهزة، الشبكات، وما إلى ذلك).

وفي الوقت نفسه، عملت قواعد البيانات على فصل عملية تحليل البيانات عن البيانات نفسـها من خلال تمكين إجراء اسـتعلامات وحسـابات معقدة دون الحاجة إلى إجراء هذه التحليـلات التـي تتطلب الاطلاع والعمل مع البيانات نفسـها أو حتى فهـم الكيفية التي تـمّ مـن خلالها تجميع البيانـات وتنظيمهـا (Gitelman and Jackson 2013). وقد تم تعزيز هـذا الفصل من خلال آليات مثـل التوحيد العيـاري (Standardization) للصيغ والبيانـات الوصفيـة والتعامل مع البيانات دون النظر إلى السـياق الواردة فيه وبعيداً عن



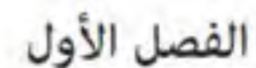
التأثـير السـياسي (Wilson 2011). والأهم من ذلك، فإن هذا الفصـل والتحرير للبيانات قد مكّن السـلطة / المعرفة المتعلقة بقواعد البيانات من الانتقال والانتشار وليتم توظيفها مـن قبل الآخرين بعيداً عـن الأعمال الداخلية المعقدة فيها وتاريخها وسياسـة الإنتاج لها (بالطريقة نفسـها التي يستفيد فيها السـائق من السيارة من دون معرفته كيفية صنع كل أنظمتهـا المعقدة أو ماذا تعمل أو كيفية تفاعلها لتشـكيل تجربة القيادة). ومن ثم يمكن للباحثين الاستفادة من قواعد البيانات الحكومية مثل التعداد السكاني أو المسح التجاري أو المؤشرات الاقتصادية دون معرفة الدوافع السياسـية لِـمَ وكيف تمّ بناء قواعد البيانات، أو الجوانب الفنية لإنتاجها، أو وجود الألفة الشخصية مع الظواهر الملتقطة. على سبيل المثال، اسـتخدام الوزارة الإيرلندية للبيئة لقواعد بيانات عقارية غير مكتملة في آيرلندا في مرحلة ما بعد الانهيار الاقتصادي 2008 (متوفرة على الرابط:

(http://www.housing.ie/OurServices/Unfinished-Housing-Developments.aspx

حيـث كان للمرء أن يجري البحث، والتصور، والاسـتخلاص للنتائـج حول العقارات في إيرلندا دون أي معرفة مسبقة عن تاريخ وسياسات المسح، وكيف تمّ القيام به، ودون زيارة أي من تلك العقارات (انظر Kitchin et al., 2012a, b). فمثل هذا التحرير والفصل مكّن مـن العمل مع قواعد البيانات على أنها غير متغيرة ومسـتقرة وقابلـة للتحول والنقل عبر المكان والزمان (Latour 1989).

وتسـتضيف البنى التحتيـة للبيانات قواعد البيانـات وتصل بينها لتشـكّل بنية تقنية اجتماعيـة أكثر تعقيداً. وكما هـو الحال مع قواعد البيانات، لا يوجد شيء كامن أو مُعطى حول كيفية تكوين هذه الهياكل الأرشـيفية والتشـاركية. وفي الواقع، وكما نوقش في جميع أنحاء الكتاب، تعرّض تصميم وإدارة البنى التحتية للبيانات إلى تحديات تقنية وسياسية تم التصدي لها من خلال نقاشـات فوضوية ومفاوضات عبثية تمت في سـياق أتصف بالعديد مـن الأجندات (جداول الأعمال) والقولبـة الحكومية. فالحلول التي تمّ إيجادها والمتعلقة بالمعايـير، والبروتوكولات، والسياسـات، والقوانين كانت بطبيعتهـا ذات آثار تقريبية كونها هدفت إلى تثبيت أرضية مشـتركة وتعميم الممارسـات ونشرها بين المطورين والمستخدمين (Lauriault 2012)، والتمويـه عـلى وتخفيف حدة التوتر والتباين بين قابلية الاسـتخدام وإتاحته والحدّ من التخصيص وتقييد الابتكار، ومنع الطرق البديلة لهيكلة وترتيب البيانات





(Star and Ruhleder 1996). وبالنظـر إلى حدة التوتـرات، ينبغي التأكيد على عمليات التقريب باستمرار وبشكل متكرر من خلال التطبيق، والإدارة، وأنظمة الحوكمة (Star and Lampland, 2009). لذلك، أكّد Star and Ruhleder (1996: 112) بأنه «لا وجود لمركز مطلـق يصدر منه التحكم والمعايير؛ وكذلك، لا أطـراف مطلقة ... مع كون البنية التحتية شيء ظاهر للناس في الممارسة العملية، ومتصلاً بالأنشطة والهياكل».

فهذا الظهور، وعلى الرغم من عدم تركزه بشـكل كامل، ليس شـكلاً حراً وهو يتشـكّل من خلال علاقات بنيوية واسـعة، فقد جـادل Graham and Marvin (2001) بأن البنى التحتية هي تأسيسـية نتيجة «تراكمات بعيدة المدى من التمويل، والتكنولوجيا، والدراية، والسلطة التنظيمية والجيوسياسية» (ص. 12) والمحافظة على «الهندسة الفنية الاجتماعية للسـلطة» (ص. 11) من المصالح الاجتماعية المتحجرة. وتشمل هذه التراكمات أنظمة من التشريعات تسـعى من الناحية القانونية ومن خلال أشـكال الحوكمة لتحديد كيفية إدارة وتحليل ومشـاركة البيانات، على سبيل المثال قوانين حماية البيانات (انظر الفصل العاشر). ولذلك، اقترح Starr (1987: 8) أن البنية التحتية للبيانات لديها:

نوعـان من الهياكل التنظيمية – الاجتماعية والمعرفية: فالتنظيم الاجتماعي يتكون من العلاقـات الاجتماعية والاقتصادية لأفراد العينة الممثلة، وأجهزة الدولة، والشركات الخاصة، والمنظمات المهنية والدولية، وغيرهم من المشاركين في إنتاج تدفقات البيانات من مصادرها الأصلية وحتى نقاط التحليل، والتوزيع، والاسـتخدام، في حين يشـير التنظيم المعرفي لهيكلة البيانـات نفسـها، بما في ذلك حدود الاسـتعلام، والافتراض عن الواقـع الاجتماعي، وأنظمة التصنيف، وطرق القياس، والقواعد الرسمية لتفسير وعرض البيانات.

كما أكّد Dourish and Bell (2007)، أنه لا يمكن النظر بصورة صرفة لقواعد البيانات والبنى التحتية كوسـائل وأدوات ثانوية لأنها تشـمل بطبيعتها نواحي ثقافية، واقتصادية، ومعرفية، وهي غارقة في الأهمية على المسـتوى الاجتماعي. ولذلك اقترحا منظورين يمكن مـن خلالها فهم البنى التحتيـة للبيانات: المنظور الأول يتضمن قراءة اجتماعية سياسـية يُدرس من خلالها «بلورة العلاقات المؤسسـية» (ص. 416). أما المنظور الثاني فهو القراءة التجريبية التي تدرس " كيفية تشـكيل البنى التحتية للبيانات لإجراءات الأفراد وخبراتهم"



(ص. ٤١٧). وفي كلتا الحالتين، تُفهم البنى التحتية للبيانات ككيانات مترابطة. ويُعيد هذا الترابط تشـكيل العالم بشكل عرضي بناء عليها، وهي بدورها تتشكّل من قبل العالم كذلك. ولكي نصل إلى اسـتخدام قواعد البيانـات والبنى التحتية للبيانـات والاعتماد عليها لفهم العالم والقيام بأعمال فيه، فإن ممارساتنا المنطقية والمادية تتكيّف وتتمحور كاستجابة لهما (Star and Ruhleder 1996). ولا ينعكس العالم في البيانات فقط، بل يتغير بها؛ «فالعمل عـلى إنتاج وحفظ وتبادل البيانات ومشـاركتها يقوم بإعادة تشـكيل العـوالم التنظيمية، والتكنولوجية، والثقافية من حولهم» (Ribes and Jackson 2013: 147).

وبعبـارة أخرى، لا تدعـم قواعد البيانـات والبنى التحتية للبيانات البحوث ببسـاطة، ولكنها تُغير وبشكل جذري الممارسات والنظم البحثية - الأسئلة المطروحة، وكيفية طرحها، وطريقة الإجابـة عليها، وكيفية تطبيق الإجابات، ومن يجري البحث وكيف يعملون بصفة باحثـين (انظر الفصل الثامن). فعلى سـبيل المثال، وفي دراسـتها لتطور التعداد السـكاني الكنـدي وأطلس كندا، بيّنت Lauriault (2012) بالتفصيل كيف تطور كل تعداد بشـكل متواتر ومتكرر على أسـاس نماذج للعالم شكلت طُرقاً لتصور وعرض كندا. وقد جادلت بأن محفوظات البيانات والبيانات نفسـها تشـكّل «نظام ذاكرة خارج الجسم المؤسسي يسمح بسرد قصص عن طبيعة كندا... من خلال الخرائط والرسـوم البيانية والنماذج والإحصاءات التـي تعتمـد على أجهزة الاستشـعار، والبيانات، وقابلية التشـغيل البيني، ومعايير رسـم الخرائط على شـبكة الإنترنت، والبوابات الإلكترونية، والبيانات الوصفية والنماذج، والعلوم، والهياكل المفتوحة» (ص. ٢٧). في المقابل، تعدّل هذه القصص النماذج الأساسـية ومن ثم البنية التحتية للبيانات، مما يعمل على تحوير الوسائل التي يتم من خلالها إنشاء القصص.

ويتطلب تفهّم قواعد البيانات والبنى التحتية للبيانات التفكيك والتفريغ بعناية لطبيعة نشوئها، وتداخلها، وعلائقيتها، وسياقيتها (Star and Ruhleder 1996). وهذا يعني البحث فيـما وصفـه Bowker and Star (1999: 34) بأنه انعكاس الأسـاس (Infrastructural Inversion) أي أن نُدرك أعماق الترابط بين الشـبكات التقنيـة والمعايير الفنية، من جهة، والعمل الحقيقي للسياسة وإنتاج المعرفة من جهة أخرى. وكما بيّنت Lauriault (2012)، فـإن هذا يتطلب أيضاً تحليلاً تاريخيـاً يوثّق كيف تتطور قواعد البيانـات والبنى التحتية للبيانات مع مرور الوقت وتغير المكان.



مجاميع البيانات وثورة البيانات (Data Assemblages and the Data Revolution) :

كان الجدل والحجة الرئيسية المعروضة في هذا الفصل بأن التفكير بالبيانات لا يكون بالطريقة المباشرة، ولا يمكن أن توجد البيانات بشكل مستقل عن الأفكار، والآليات، والتقنيات، والنظم، والناس، والسياقات، بغض النظر عن كونها في كثير من الأحيان يتم تصورها بهذه الطريقة (Lauriault 2012; Ribes and Jackson 2013). فالبيانات يتم إنشاؤها بوصفها نتاجاً لكثير من العقول العاملة في مختلف الحالات والأوضاع، وهي مؤطرة وتتشكل في ظروف وهياكل مختلفة.

ولذا فإن من طرق فهم البيانات هو التفكير بها كقضية مركزية لتجمّع فني اجتماعي معقد، ويتكوّن هذا التجمّع للبيانات من العديد من الأدوات والعناصر التي تزداد تشابكاً، وتطوراً، وتتحوّر مع مرور الزمن وتبدّل المكان (انظر الجدول3-1). وتصوغ كل أداة مصحوبة بعناصرها ما هو ممكن، وما هو مطلوب، وما هو متوقع من البيانات. وعلاوة على ذلك، فهي تتفاعل بعضها مع بعض ويشكّل بعضها بعضاً من خلال شبكة معقدة ومتغيرة من العلاقات المتعددة الأوجه (انظر الشكل 3-1). وكما ادّعى كل من Ribes and Jackson (2013)، فإن هذه الأدوات لا تقوم بتأطير ماهية وكيفية إنتاج البيانات وإلى أيّ حدّ من الممكن توظيفها فحسب، ولكنها بحدّ ذاتها منظمة ومُدارة لتقوم بإنتاج مثل هذه البيانات. وعلى ذلك، فإن البيانات ومجاميعها تتشارك التحديد (Co-determinous) وتتبادل عملية التشكيل كلّ منها للآخر(Mutually Constituted)، وترتبط ببعضها من خلال مجموعة من الممارسات والعلاقات المنطقية والمادية الطارئة، والعلائقية، والسياقية. وعلى ذلك، يختلف كل تجميع للبيانات في نسق ترتيب العناصر وسياقها، ولكنها تتشارك في قواسم مشتركة وتأثير كلّ منها في الآخر بسبب اعتماد الأدوات على بعضها وتداخلها والاتفاقيات التي تغطي مجاميع البيانات، ومع نشوء أفكار ومعارف جديدة، واختراع التقنيات، وتطوير المهارات، وانفتاح الأسواق، فإن مجاميع البيانات تتطور، وتتحوّر، وتندمج، وتتفكك. ونتيجة لذلك، هناك تنوع كبير من تجمّعات البيانات عبر النطاقات والاختصاصات المختلفة.

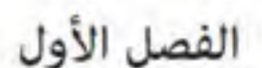


جدول رقم (٣-١)

الأدوات والعناصر لمجاميع البيانات

الجهاز / النظام	العناصر
نظم الفكر	طرائق التفكير، الفلسفيات، النظريات، النماذج، الإيديولوجيات، المنطقيات،... إلخ.
أشكال المعرفة	النصوص البحثية، الأدلة، المجلات، المواقع الإلكترونية، الخبرة، الخُطب، منتديات الدردشة،... إلخ.
التمويل	نماذج الأعمال، الاستثمار، رأس المال الاستثماري، المنح، التبرعات والصدقات، الربح،... إلخ.
الاقتصاد السياسي	السياسة، الأنظمة الضريبية، الرأي العام والرأي السياسي، الاعتبارات الأخلاقية،... إلخ.
الجوانب الإدارية والقانونية	معايير البيانات، صيغ الملفات، متطلبات النظم، اللوائح البروتوكولية، القوانين، الترخيص، أنظمة الملكية الفكرية،... إلخ.
الأصول المادية والبنى التحتية	الأوراق / الأقلام، الحاسبات الآلية، الخدمات الرقمية، أجهزة الاستشعار، الماسحات الضوئية، قواعد البيانات، شبكات الاتصال، الخوادم،... إلخ
الممارسات	التقنيات، طرق التنفيذ، السلوكيات المستفادة، الاتفاقيات العلمية،... إلخ.
المنظمات والمؤسسات	سجلات الأرشفة، الشركات، المستشارون، المصنّعون، بائعو التجزئة، الجهات الحكومية، الجامعات، المؤتمرات، النوادي والمجتمعات، واللجان والمجالس، المجتمعات الممارسة،... إلخ.
الشخصيات والمجتمعات	منتجو البيانات، القيّمون على البيانات، المديرون، المحلّلون، العلماء، السياسيون، المستخدمون، المواطنون،... إلخ.
الأماكن	المختبرات، المكاتب، مواقع العمل، مراكز البيانات، مجموعات الخوادم، مجمعات الأعمال،... والقوى البشرية المتصلة بها.
الأسواق	للبيانات، ولمشتقاتها (على سبيل المثال، النصوص، الجداول، المخططات، الخرائط)، والمحلّلون، وبرمجيات التحليل، والتفسيرات،... إلخ.





فهذا الكتاب يبحث في تجمعات البيانات الناشئة، والمتطورة، المنتجة للبيانات المفتوحة، والبنى التحتية للبيانات، والبيانات الكبيرة، وللقيام بذلك يطرح الكتاب ثلاث مسائل رئيسية للمناقشـة: أولاً، هناك حاجة لتطوير طرق مبدئية وفلسـفية لفهم البيانات. فهناك نقص ملحوظ في التفكير النقدي والبحثي عن البيانات وفي البيانات بحدّ ذاتها من حيث التكوين والتشـغيل للمجاميع المحيطة بها، وعلى الخصوص بالمقارنة مـع التركيز والاهتمام اللذين تحظى به مفاهيم المعلومات والمعرفة. وذلك على الرغم من أن البيانات هي عنصر أساسي في إنتاج المعرفة. ثانياً، هناك ثورة بيانات جارية تعمل على تشكيل لحظة حاسمة في تطور وتحوّر تجمعات البيانات. ونظراً لالتقاء العديد من التطورات الناشـئة في مجال الحوسبة، والتقنيات المنهجية، والمجالين السياسي والاقتصادي، فإن حجم البيانات، وتنوعها، وسرعتها، ودقتهـا، وتوافرها، والكيفية التـي يتم بها معالجة البيانات وتحليلهـا وتخزينها وتوظيفها لزيـادة التبصر وتحصيل القيمة، قد شـهد تحـولاً جذرياً. ثالثاً، نظـراً لمختلف التحديات التقنيـة والأخلاقية والعلمية التي أثارتها ثورة البيانات، فإن هناك حاجة ملحّة لتطوير فهم مفصل لتجمعات البيانات الجديدة والناشئة التي يتم إنشاؤها، ولذا تهدف الفصول العشرة التالية إلى توفير نظرة إجمالية على نطاق واسع، وشامل، ودقيق لهذه التجمعات وتسليط الضوء على القضايا التي تتطلب مزيداً من الاهتمام والبحث.

شكل (٣-١)

التقاطع بين أدوات مجاميع البيانات

Systems of thought
نظم الفكر
Governmentalities and legalities
الجوانب الإدارية والقانونية
Political economy
الاقتصاد السياسي
Finance
التمويل
The marketplace
الأسواق
Materialities and infrastructures
الأصول المادية والبنى التحتية
Organisations and institutions
المنظمات والمؤسسات
Subjectivities and communities
الشخصيات والمجتمعات
Places
الأماكن
Practices
الممارسات
Forms of knowledge
أشكال المعرفة


المكتبة الاقتصادية
قناة المكتبة على التيليجرام
ECONLIBRARY





الفصل الثاني

البيانات الصغيرة (Small Data)، والبنى التحتية للبيانات (Data Infrastructures)، ووسطاء البيانات (Data Brokers)

قبل بضع سـنوات مضت كان مصطلح «البيانات الصغيرة» نادر الاسـتخدام، إن لم يكن معدوماً. وقد ذاع انتشاره مصطلحاً بحكم الواقع على النقيض مما سمّي «البيانات الكبيرة». فـكلّ البيانات كانـت في الواقع بيانات صغيرة ومن ثم لم تتطلـب تمييزها بهذا الإصطلاح. وسـيجري لاحقاً مناقشـة ما قام على تأسـيس البيانات الكبيرة بالتفصيل في الفصل الرابع، ولكن من الضروري تسليط الضوء على الفرق بين البيانات الصغيرة والبيانات الكبيرة، وذلك قبـل التوجه إلى مناقشـة كيفية تحوّل البيانات الصغيرة من خـلال البنى التحتية الجديدة للبيانـات والأسـواق المقدّرة ببلايين الدولارات، كما سـيتم لاحقاً في الفصل التالي مناقشـة انفتاحها لتقبل المشاركة وإعادة الاستخدام.

فمن الناحية التقليديـة، ونظراً للتكاليف والصعوبات المصاحبـة لعمليات إنتاج ومعالجة وتحليـل وتخزين البيانات، فقد تم إنتاج البيانات الصغيرة بطرق مضبوطة بأحكام باسـتخدام تقنيات أخذ العينات التي تحدّ من نطاقها، وصلاحيتها الزمنية، وحجمها (H.J. Miller 2010). وبينـما يمكن أن تكون بعض قواعد البيانات كبيرة جـداً في الحجم، مثل قاعدة بيانات التعداد الوطني، ولجعل ممارسـة تجميع هذه البيانات قابلة لـلإدارة، فإن عملية إنتاجها عادة ما تتم خلال بضع سـنوات وتكون محدودة النطاق (في حالة التعداد، يجرى في العادة مرة واحدة كل عشر سـنوات ويكون مقتصراً على ثلاثين إلى أربعين سؤالاً فقط). وعلاوة على ذلك، فقد كانت دقة البيانات الصغيرة العامة (Small Tertiary Data) في كثير من الأحيان ضعيفة جداً (كانت، على سـبيل المثال، على مسـتوى المقاطعات أو الدول وليس الأفراد والأسر)، كما أن الأسـاليب المسـتخدمة لإنتاجها غير مرنة على الإطلاق (على سبيل المثال، من المستحيل تعديل الأسئلة أو الإضافـة لها أو الإزالة منها بمجرد التهيئة للتعداد وإعداده والبدء بإدارته لأن ذلك سـينعكس سـلباً وبشـكل جديّ على القدرة على المعالجة والتحليل). وفي المقابل، تتميز البيانات الكبيرة



بكونها تُنتج بشكل مستمر، وتسعى لأن تكون شاملة وغارقة في التفاصيل ودقيقة على مستوى النطـاق الـذي تغطيه، ومرنة وقابلة للتطويـر في عملية إنتاجها (انظر الجـدول 1-2 والفصل الرابع).

جدول رقم (١-٢)

مقارنة البيانات الصغيرة والبيانات الكبيرة

	البيانات الصغيرة	البيانات الكبيرة
الحجم	محدودة	ضخمة جداً
درجة التفصيل / الشمولية	عينات	كامل مجتمع الدراسة
الدقة والتحديد	من تقريبية وضعيفة إلى مترابطة وقوية	محكمة وقوية
المنطقية / العلائقية	من ضعيفة إلى قوية	قوية
السرعة	بطيئة، مجمدة التأطير/ محزّمة	سريعة، مستمرة
التنوع	من محدودة إلى واسعة	واسعة
المرونة وقابلية التوسع	منخفضة إلى متوسطة	عالية

وفي الوقـت الذي تعد فيه البيانـات الصغيرة واحات صغيرة ضمن صحـارى البيانات، تنتج البيانات الكبيرة طوفان بيانات حقيقي. وقد أدّى ذلك بالبعض إلى التسـاؤل عما إذا كانت البيانات الكبيرة قد تؤدي إلى زوال البيانات الصغيرة أو تضاؤل مكانة الدراسات التي تعتمـد على البيانات الصغيرة نظـراً لمحدوديتها في الحجم، والنطـاق، والصلاحية الزمانية. فقد لاحظ Sawyer (2008) أن وكالات تمويل البحوث الأكاديمية، على سـبيل المثال، توجه مـوارد التمويل المحدودة الخاصة بها إلى مجالات البيانـات الغنية على نحو متزايد، وذلك ربما بسـبب التداخل بين حجم البيانـات وسرعتها مع مقدار التبـصر والتوظيف والقيمة، وأن هذه الوكالات تحول الأسـئلة البحثية التي من الصعـب أن تنتج بيانات كبيرة لجهات التمويل الأخرى المتعددة وتمنحها وضع هامشي داخل وخارج هذه المؤسسـات الأكاديمية. وبعيداً عن البحوث الأساسية، فقد استكمل هذا التحرك من خلال توجه العديد من وكالات التمويل نحو البحوث التجريبية التطبيقية ذات الشراكة الصناعية. وقد أدى إعادة الترتيب



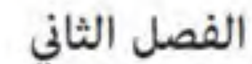
لأولويات البحث إلى سوء الفهم لطبيعة البيانات الكبيرة وقيمة البيانات الصغيرة، فالبيانات الكبيرة ربما تهدف إلى أن تكون شاملة، ولكن كما هو الحال مع جميع البيانات فإن كليهما عبارة عن تمثيل (Representation) وعينة (Sample)، فماهية البيانات التي يتمّ التقاطها تتشكّل في الواقع من خلال التالي:

- مجـال إطار عـرض/ أخذ العينات (حيث يتـم توظيف أجهزة التقـاط البيانات وماهية إعدادات التهيئة لها/ ومعاملاتها؛ مثل من يستخدم المكان أو وسائل الإعلام، وعلى سبيل المثال من الذي ينتمي إلى وسـيلة التواصل الاجتماعي الفيسـبوك أو يتسـوّق في محلات وول مارت،..إلخ).
- التقنيـة والمنصة المسـتخدمة (المسـوحات المختلفة، وأجهزة الاستشـعار، والعدسـات، والمطالبـات النصية، والمخططات والتصاميم،.. إلخ، وكلها تنتج فروقاً وتباينات في ماهية البيانات التي يتم إنشاؤها).
- السياق الذي يتمّ إنشاء البيانات فيه (الأحداث الجارية تعنى أن البيانات تقع دائماً نسبة إلى الظرف الذي تقع فيه).
- تبويبات (أنطولوجية) البيانات المستخدمة (كيفية معايرة البيانات وتصنيفها).
- البيئة التنظيمية فيما يتعلق بالخصوصية، وحماية وأمن البيانات.

وتلتقط البيانات الكبيرة في العموم ما هو سـهل التقاطه – أي البيانات التي يتم التعبير عنها علناً (ما كُتب، ومُرّر، ومُسح ضوئياً، وما تمّ استشعاره،.. إلخ؛ وتصرفات الناس وسلوكهم؛ وحركة الأشياء) – وكذلك البيانات المستنفذة والتي هي ناتج وظيفة رئيسية وليست المخرج الرئيسي، والبيانات الوصفية التي تمّ إنشاؤها تلقائياً. وتأخذ البيانات الكبيرة هذه البيانات بالقيمة الاسـمية، على الرغم من أنها قد لا تكون قد تم تصميمها للردّ على أسئلة محددة والبيانات الناتجة قد تكون عشوائية وغير دقيقة (انظر الفصل التاسع).

أما البيانات الصغـيرة فقد تكون محدودة في الحجم والسرعة، ولكن لديها تاريخاً طويلاً من التطوير، مع منهجيات مقررة وأطوار للتحليل، وسجل حافل من الإجابات الناتجة للأسئلة العلمية. ويمكن لدراسـات البيانات الصغيرة أن تكون مصممة بشكل أفضل بكثير للرد على





أسئلة بحثية محددة والاستكشـاف المفصّل والمعمّق للطرق المتنوعة، والسياقية، والمنطقية وغير المنطقيـة التي يتعامل بها الناس، ويتمّ فهم العالم بها، وكيفية عمل العمليات العلمية (انظر الفصلين الثامن والتاسع). ويمكن أن تركز البيانات الصغيرة على حالات محددة وتُخبر عن قصص فردية، وسياقية، ودقيقة. وبسبب أنها تستهدف ظاهرة معينة، فهي تركز بشكل مكثـف على قضية وعادة ما تكون أقل حاجة للموارد بشـكل مكثف: وتسـعى دراسـات البيانـات الصغـيرة على التنقيب عن الذهب من خلال العمل في شـقّ ضيق للغاية، في حين تسـعى دراسـات البيانات الكبيرة إلى استخراج شذرات من خلال التعدين في حفرة مفتوحة وشفط وغربلة مساحات شاسعة من الأراضي. ومن ثم فقد برهن Sawyer (2008) على أن التمويل يحتاج إلى اسـتهداف الإجابة عن الأسـئلة الحرجة، سواء كانت معالجتها باستخدام بيانات صغيرة أو كبيرة، وليس مجرد أن توجّه نحو المشاريع التي يمكنها الوصول إلى كميات هائلة من البيانات على أمل أنها تحوي بطبيعتها رؤى وأفكاراً مفيدة.

وهكـذا، يمكن القول إنه من الممكـن تحصيل المزيد من القيمة والـرؤى من البيانات الصغـيرة من خلال توسـيع نطاقهـا إلى مجموعات بيانـات أكبر، وجعلهـا متاحة لإعادة الاسـتخدام من خلال البنـى التحتية للبيانات الرقمية والتي تجعلهـا إضافة إلى ذلك قابلة لمختلف أشكال التحليل الحسابي، والحفاظ عليها للأجيال القادمة. ونتيجة لذلك، وكما هو الحال مع البيانات الكبيرة، فقد استثمرت موارد كبيرة في السنوات الأخيرة لإنشاء مثل هذه البنى التحتية للبيانات، والتي أحدثت تغييرات كبيرة في مشهد البيانات.

حاويات البيانات (Data Holdings)، محفوظات البيانات (Data Archives)، والبنى التحتية للبيانات (Data Infrastructures):

تـمّ جمع البيانات معاً وتخزينهـا لفترة طويلة من التاريخ المسـجّل، وقد كانت هذه الممارسـات غير رسمية ورسمية على حدّ سواء في طبيعتها، فالممارسات غير الرسمية تتكوّن ببسـاطة من جمع البيانات وتخزينها، في حين تتكوّن الممارسـات الرسـمية من مجموعة الممارسـات الفاعلة ذات العلاقة بإدارة دورة حيـاة البيانات (Data Curation) والهياكل المؤسسية المصممّة لضمان الحفاظ على البيانات من أجل الأجيال القادمة. وقد يكون من الأفضل وصف الممارسات غير الرسمية بأنها حاويات البيانات، في حين أن الممارسات الرسمية

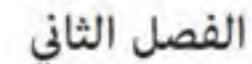


هي محفوظات (أرشيفات) البيانات. ويجري، في كثير من الأحيان، الخلط بينهما أو تستخدم بالتبادل، جنبا إلى جنب مع غيرها من المصطلحات مثل النسخ الاحتياطي (Backup)، وبنك المعلومات (Information Bank)، ومستودع البيانات (Data Repository)، ولكن المحفوظات الأرشيفية تمتلك خصائص مميزة جداً عند أمناء المحفوظات (Archivists) المهنيين.

فالمحفوظات هي مجموعات رسمية منظمة من البيانات التي يتم هيكلتها وتوثيقها ونشرها بشكل فعال، مصحوبة بالبيانات الوصفية المناسبة لها، مع التخطيط الفعّال لقضايا المحافظة عليها وإتاحة الوصول إليها وإمكانات اكتشافها (Lauriault et al. 2013). فالمحفوظات تهدف بشكل واضح وصريح وطويل الأمد، للحفاظ على البيانات لإعادة استخدامها في المستقبل. إن مفهوم الحفاظ، في هذا السياق، لديه استخدام خاص محدّد على أنه مجموعة من الأنشطة المُدارة والمدعمّة «بالمبادئ، السياسات، القواعد، والإستراتيجيات الرامية إلى إطالة أمد الكائن (البيانات) من خلال المحافظة عليه في حالة مناسبة للاستخدام، سواء على شكله الأصلي أم في شكل مُستدام أكثر ثباتاً، مع الإبقاء على الشكل الفكري للكائن» (InterPARES 2, 2013). ويسعى الأرشيف للحفاظ على مجموعة سجلات كاملة، وليس مجرد البيانات؛ أي أنها تشمل كل الوثائق الداعمة والبيانات الوصفية وغيرها من المواد ذات الصلة التي تحتوي على تفاصيل المنشأ والسياق فيما يتعلق بكيفية إنتاج البيانات والطريقة التي ينبغي أن تعامل وتحلّل وتفسّر بها. فالنهج المتبع في الحفاظ على البيانات مبني على الإدراك التام بأن التقنيات والبروتوكولات والمبادئ التوجيهية لأفضل الممارسات هي عرضة للتغيير والتقادم، بمعنى أن البيانات سوف تحتاج للتنقل والترحيل عبر منصات وتقنيات بسبب الابتكارات الجديدة التي تأتي في السياق، كما يُدرك النهج المتبع بأنه دون ممارسات فعالة لإتاحة وإعادة الاستخدام والمشاركة للبيانات فإن البيانات معرضّة للتلف، أو الضياع، أو الحرمان من بياناتها الوصفية السياقية والوثائق الداعمة (Borgman 2007; Dasish 2012). ويسعى المحفوظ الأرشيفي إلى توفير موارد جديرة بالثقة، من خلال الحفاظ على سلامة البيانات (Data Integrity) مع مرور الوقت.

وفي بعض الحالات، قد يُلزم مؤتمنو البيانات (Data Custodians) بمسؤولية قانونية لحفظ وأرشفة البيانات بشكل مناسب لإعادة الاستخدام الحالي والمستقبلي، كما قد يكون





المؤتمن على الأرشفة والقائم عليها نفسه سلطة قانونية تشريعية. فعلى سبيل المثال، فإن المحفوظات الأرشيفية الوطنية والوكالات الإحصائية الوطنية هي الكيانات القانونية التي عُهد إليها جمع وإنتاج وأرشفة قواعد بيانات محددة مثل السجلات الشخصية للأفراد والحسابات القومية، والتعدادات والمسوح السكانية. وينبغي أن تمتثل المحفوظات أيضاً للتشريعات المتعلقة بحقوق الوصول إلى البيانات، والخصوصية، والمبادئ الأخلاقية، وحقوق الطبع والنشر، والملكية الفكرية، وتحديد المسموح لهم للوصول إلى البيانات وما يمكنهم القيام به من خلال تحديد الصلاحيات على البيانات. على سبيل المثال، في العديد من البلدان، من الممكن نشر سجلات التعداد في شكل إجمالي، مع تقييد نشر السجلات الشخصية إلى ما بعد مئة عام، فالالتزام بإتاحة الوصول وإعادة الاستخدام للبيانات لا يستلزم بالضرورة العمل على نموذج مفتوح (انظر الفصل الثالث)، ولكن ينبغي أن يتم ضمان توافر البيانات في المستقبل للمجتمع المستهدف بطريقة قانونية (CCSDS, 2012). إن المجتمع المستهدف هو عادة مجموعة معينة من العلماء أو المتخصصين في مجال السياسات الذين لديهم في العادة احتياجات محددة تماماً من البيانات، ومتطلبات وظيفية ودعم مما يعني أن حلول الأرشفة العامة ستكون ذات فائدة محدودة (Lauriault et al. 2013).

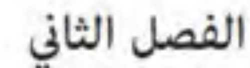
ومن ثم، فإن محفوظات البيانات ليست مجرد مخازن بيانات أو أنظمة نسخ احتياطي (والتي تحفظ البيانات في حال حدوث أعطال على المدى القصير مثل عطل الوسط التخزيني أو الفشل التقني)، ولكنها تُخطّط، وتنسّق، وتُدار، وتُوظّف بفعالية من قبل موظفين متفرغين ومتخصصين يعملون على إضافة القيمة وضمان الاستمرارية. فهؤلاء الأفراد، إضافة إلى الكيان الإداري الوظيفي، مسؤولون عن الأرشفة إضافة إلى ثلاث مجموعات أخرى من الأطراف الفاعلة في عملية الأرشفة وهم: منتجو البيانات، ومستهلكو البيانات، والإدارة (CCSDS 2012). فالمنتجون هم منشئو البيانات أو الكيان القانوني المسؤول عن إنتاجها؛ أما المستهلكون فهم مستخدمو المحفوظات، وأصحاب المصلحة المعنيون بها أو مجتمع معين؛ وأما الإدارة فهي الهيئة التي تشرف على العملية، وتوفّر للأرشيف ميثاقاً ونطاقاً، ولكنها لا تأخذ دوراً نشطاً في عمليات الأرشفة اليومية (CCSDS 2012; Lauriault et al. 2013). وتشكّل هذه الدوائر مجتمعة كيفية بناء وتشغيل وتطوير الأرشيف مع مرور الوقت.



في المقابــل، فإن حاويات البيانات هو مجموعة غير رســمية وغير نظامية من البيانات. وقد يعتد بعض منتجي البيانات أن النسخ الاحتياطي أو أنظمة التخزين تشكّل محفوظات أرشــيفية بحكم الواقع، ولكنها تفتقر إلى الجهاز المؤســسي، والتنظيم، والسياسات الرسمية النظامية (بالنســبة إلى المعايير، وإتاحة الوصول إليها، وإمكانية إعادة استخدامها، وحقوق الملكية الفكرية، والمعايير الأخلاقية، وإلى آخره)، وما يطمح إلى تحقيقه الأرشيف (Lauriault et al. 2013). فقــد تكون حاويات البيانات منتظمة ومُوجّهة، ولكن بدون قواعد صارمة أو أي خطط طويلة المدى تتعلّق بالمحافظة على البيانات وحفظها، ودون أي التزام لمشاركة البيانات وتبادلها أو جعلها متاحة لإعادة الاستخدام. وعلى الرغم من أن حاويات البيانات قــد تلتزم بشــكل جيد بمعايــير البيانات المتوقعــة، إلا أن هذا الالتزام يفتقــر إلى الاعتماد والمصادقة عليه ويفتقر إلى الوثائق الأخرى والبيانات الوصفية المناسبة.

وقد فُقدت معظم البيانات التي تمّ إنتاجها على مر التاريخ أو دُمّرت بســبب تخزينها في حاويات بيانات، وليس أرشــيفاً، أو لأنه تقرر الحفاظ على المعلومات المســتقاة من هذه البيانات (مثل المقالات والكتب) التي كانت تعد أكثر قيمة، وتخزينها في المكتبات. وبشكل عــام، تم الإبقاء فقط على مجموعــات البيانات الأكثر قيمة، مثل تلك المرتبطة بالمســاعي العلمية والثقافية الأساسية والسجلات الحكومية، والمعاملات الاقتصادية، والعقود القانونية. وفي حالات النماذج، الفنون، الكائنات ....إلخ والتي يمكن اعتبارها صوراً معينة من البيانات فقد تمّ تنسيقها وتخزينها بالأساس في المتاحف وصالات العرض جنباً إلى جنب مع البيانات الوصفية المرتبطة بها. وقد تمّ تخزين البيانات الخاصة بمعظم العلماء بشــكل غير رســمي عــلى ملفات وصناديق أو على مختلف محركات الأقراص الصلبة في مكاتبهم أو في منازلهم. وعندمــا يتقاعدون خارج الخدمة أو يتوفون يتمّ تدمــير معظم آثارهم، ومعهم أيّ بيانات قاموا بإنتاجها. كما أن الغالبية العظمى من البيانات التي تمّ إنتاجها لأطروحات الدكتوراه تفقد بعد الانتهاء منها، وغالباً لا يتم نشر حتى المعلومات المســتقاة من البحوث. وعلاوة عــلى ذلك، فإن مُمــوّلي البحوث تاريخياً لم يطالبوا المشــاريع التي قامــوا بتمويلها بحفظ البيانــات وتخزينها، أو إن طلبوا ذلك فقد كان لفــترة قصيرة من الزمن. وكما هو الحال في العديد من المشــاريع التي يموّلها الاتحاد الأوروبي، يُشــترط الاحتفاظ بجميع مواد المشروع بما في ذلك البيانات لمدة ســبع ســنوات، ولكن يمكن بعد ذلك التخلص منها، وهو ما يتم في





كثير من الحالات، بسبب تكاليف التخزين والظن الغالب بأن المقالات الورقية الناتجة تمثّل قيمة أكبر.

ومـع تطور حلول التخزين الرقمي، التي قلّلت التكلفة والمسـاحة التخزينية للاحتفاظ بالبيانـات، فقد بدأ التخزين الطويل الأمد على نطاق واسـع للبيانات الاعتيادية المنخفضة التكلفـة (انظر الفصل الخامس). فقـد ظهرت أولى قواعد البيانـات الرقمية التي ضمّت مجموعات البيانات في الخمسـينيات والسـتينيات من القرن الماضي وكانت تنسّق البيانات في هيـاكل مسـطّحة أو هرمية منظمـة (Driscoll 2012). ومع ذلـك، فقد كانت هذه الترتيبات غير فعالة للبحث والاستعلام عن البيانات، كما تمّ تطوير قواعد البيانات العلائقية (Relational databases) ابتـداء في عـام ١٩٧٠ من قبل إدغار كود (Edgar Codd) وقد نظّم خصائص البيانات المميزة (Attribute) في مجموعة من الجداول المترابطة، حيث يتشارك كل جدول بحقل بيانات واحد على الأقل مع جدول آخر، فعلى سـبيل المثال، قد يكون رقم الضمان الاجتماعي هو الحقل المشترك بين جدولين (المفتاح التأشيري- Indexical Key) أحدهما يحتوي على التفاصيل الشخصية (مثل العمر، الجنس، الحالة الاجتماعية، والعنوان) ويحتوي الجدول الآخر على المتعلقات المالية، ويسـمح رقم الضمان الاجتماعي بربط البيانات للاسـتعلام في كلا الجدولين معاً وإنتاج بيانات جديدة مشـتقة منهما. وإذا كان المفتاح المشـترك عنواناً أو تاريخـاً، فيمكـن عندها وضع البيانـات جغرافياً أو زمانياً. لقـد مكّنت قواعد البيانات العلائقية تنظيماً واسـتعلاماً أكـثر فعالية وتعقيداً عن البيانات الهيكلية باسـتخدام لغات الاسـتعلام الهيكليـة (SQL). وجنباً إلى جنب مع قواعد البيانات العلائقية، سـمح تطوير البيانات المجدولة (Spreadsheets) بتنظيم وتخزين كميات كبيرة من البيانات الحسـابية إضافـة لتطبيق الصيغ والمعادلات عليها مما يؤدي إلى إنتاج بيانات مسـتقاة جديدة. وقد اسـتكملت هذه الأشـكال من قواعد البيانات، في السـنوات الأخيرة، بقواعد البيانات غير الهيكلية (بلا اسـتعلام noSQL-) التي يمكنها التعامل مـع كميات كبيرة من البيانات غير الهيكلية (انظر الفصل الخامس).

ويكون إنشـاء ونمو مجموعات البيانات سهلاً في الحالات التي تنتج فيها البيانات ابتداء بصورة رقمية (نشـأت بالكود الثنـائي-binary code)، وأحياناً بطرق تلقائية ومؤتمتة، كما

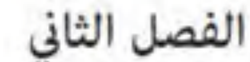


أنها عرضة للخسارة بشكل كبير بسبب التقادم في البرمجيات والمعدات. وقد بيّن O'Carroll and Webb (2012) أنه: «في حين يمكن لأي أحد التقاط كتاب كُتب قبل مئة سنة، والنظر فيه، وقراءة صفحة منه، فإنه لن يكون ذلك بالمثل صحيحاً لقرص مرن يحتوي على ملفات نصية لبرمجيات وورد بيرفكت أنشـئت قبل ٢٠ سـنة فقط». وفي المقابل، ينبغي أن تحوّل الأشـكال التناظرية/ التماثلية للبيانات إلى الصورة الرقمية الثنائية على سبيل المثال، مسحها ضوئياً أو إعادة تسـجيلها وإدخالهـا، والتي يمكن أن تكون مكلفة وتسـتغرق وقتاً طويلاً، عـلى الرغم من إمكانية أتمتة هذه العملية آلياً إلى حد ما. وقد حدّت التكاليف والنفقات المرتبطـة بالرقمنـة (Digitisation) من حيث التكلفة الماديـة، وقت الموظفين، والمعدات المتخصصـة، من تطبيقها في العديد من المحفوظات الأرشـيفية التناظرية التي تحتفظ بها المتاحف والمكتبات والمجموعات الشخصية الخاصة. وبينما كافحت هذه المؤسسات لتمويل أنشطة الرقمنة، فإن جهات خيرية غير ربحية مثل أرشيف الإنترنت (http://archive.org) وتجارية مثل قوقل (Google) تسـاعد على القيام بمثل هذه الأنشـطة، وذلك باسـتخدام مواردهـا الذاتية أو تبرعات المحسـنين من عامة الأفراد مما يجعلهـا متاحة مجاناً للجميع (انظر الفصل الخامس).

وفي جميع الحالات، فإن البيانات ضمن حاويات البيانات الرقمية والمحفوظات الأرشيفية يمكن مشـاركتها وإعادة استخدامها بسـهولة وبتكلفة هامشية منخفضة، وذلك على الرغم من أنه من الممكن الحدّ فيما يتعلق بالوصول وإعادة الاسـتخدام بواسطة سياسات حقوق الملكيـة الفكرية. وعلاوة على ذلك، فإنه من الممكن معالجتها وتحليلها من خلال الأدوات الحسـابية القوية، إذ أصبحت الإجراءات والحسابات التي كان من الصعب القيام بها يدوياً أو باسـتخدام التقنيات التناظرية على سـبيل المثال، المسـطرة الحاسـبة التناظرية (Slide rules) ممكنـة في غضـون بضعة أجزاء من الثانية، وأكثر في بنـاء واختبار النماذج وأنظمة المحاكاة المعقدة. كما يمكن أيضاً ربط حاويات البيانات الرقمية والمحفوظات الأرشيفية معاً بسهولة نسبياً ومدّها إلى أشكال أخرى من البنى التحتية للبيانات.

فالبنيـة التحتية للبيانات هي الوسـيلة المؤسسـية والمادية والرقميـة لتخزين وتبادل واسـتهلاك البيانات عبر التقنيات المرتبطة بعضها ببعض، وعلى مدى العقدين الماضيين على





وجــه الخصوص، فقد تــمّ بذل جهد كبير لتطوير وتعزيز هذه البنــى التحتية، والتي تأخذ عدة أشــكال: الكتالوجات، والأدلــة، والبوابات، ومراكز تبادل المعلومات، والمســتودعات، والمســتودعات الرقمية الموثوقة، والبنــى التحتية الســيبرانية (Cyber-Infrastructures) (Lauriault et al. 2007). وغالبــاً ما تســتخدم هذه المصطلحات بالتبادل ويجري الخلط بعضهــا ببعض، وذلك على الرغم من أنها تمثل أنواع مختلفــة من الكيانات (انظر الجدول 2-2 على سبيل المثال).

فالكتالوجــات، والأدلة، والبوابــات هي موارد مركزية تفصّل وتربــط حاويات البيانات والمحفوظات الأرشيفية التي تحتفظ بها المؤسسات على المستوى الفردي. وقد توفّر قوائم مفصلــة إلى حدّ ما عن مجموعات البيانات التي تضمّها، كما يمكن أن تكون بمثابة مجاميع للبيانات الوصفية، ولكن لا يســتضيف الموقع نفســه البيانات (O'Carroll et al. 2013). أما مستودع البيانات أو مركز تبادل المعلومات فهو مكان مشترك لتخزين حاويات البيانات والمحفوظات الأرشيفية والوصول إليها. وقد توفر هذه المستودعات والمراكز بعض خدمات البيانات من قبيل البحث والاســترجاع، وإدارة البيانــات ومعالجتها، ولكن عملية إنتاج كل حاوية أو أرشيف تتم بشكل مستقل، ولذا فهي ربما تفتقر إلى صيغ ومعايير وبيانات وصفية وسياسات مشتركة للبيانات. ومع ذلك، يسعى مستودع البيانات لضمان التزام كل أرشيف أو حاويــة بمجموعة من المواصفات المطلوبة ويقــوم بالمراجعة، والتدقيق، والتصديق على منح الشــهادات لضمان ســلامة البيانات وتعزيز الثقة بين المســتخدمين (Dasish 2012). وعرّف O'Carroll et al. (2013) نوعين من المســتوعات. مستودعات الموقع الواحد التي تســتضيف كافة مجموعات البيانات في موقع واحد، ويمكن الوصول إليها من خلال واجهة على شــبكة الإنترنت، على الرغم من إمكانية احتفاظها بنســخ احتياطية أو مواقع مثيلة في أماكن متعددة. والمستوعات المتعددة المواقع التي تستضيف البيانات ضمن هيكل موحد يسمح بالوصول إلى مجموعات البيانات القائمة في عدد من المواقع المؤسسية.

فالمســتوع الرقمي الموثــوق (TDR – Trusted Digital Repository) هو مســتودع مرخــص بموقــع واحد أو متعــدد المواقع تتمثــل مهمته في توفير وصول آمــن وموثوق به للبيانــات التــي يديرها على المدى البعيــد (RLG and OCLC 2002). ومن أجل التأهل

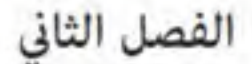
كمســتودع موثوق، فصّلــت (RLG and OCLC (2002: 5 الخصائص المطلوب توفرها في البنية التحتية لهذه المستودعات وهي كالتالي:

- تقبل المســئولية عن الصيانة طويلة الأمد للمــوارد الرقمية بالنيابة عن المودعين، ولما فيه مصلحة المستخدمين الحاليين والمستقبليين، والمستهلكين أو المجتمعات المعنية.
- امتلاك نظام مؤســسي لا يدعم فقط اســتمرارية نظام المحافظة على البيانات على المدى البعيد، ولكن يحافظ أيضاً على المعلومات الرقمية التي تقع ضمن نطاق مسؤوليته.
- توضيح المسؤولية المالية وضمان الاستدامة.
- تصميم أنظمتهــا وفقاً للاتفاقيات والمعايير المقبولة والمعتمدة لضمان الإدارة المســتمرة، والوصول، وأمن المواد المودعة داخلها.
- وضع منهجيات لتقييم المســتودع، وذلــك للتأكد من تلبيته للتوقعات المجتمع من حيث الموثوقية.
- إمكانيــة الاعتماد عــلى البنية التحتية الخاصة بها وضمان تحمّل مســؤولياتها على المدى الطويل تجاه المودعين والمستخدمين بشكل علني وصريح.
- امتلاك السياسات، والممارسات، والأداء الذي يمكن تدقيقه وقياسه.

ولكي يمكن لمســتودع مــا الحصول على هــذه المتطلبات، أكّــدت RLG and OCLC (2002: 13) أنه ينبغي أن يكون متوافقاً مع النموذج المرجعي لنظام المعلومات الأرشــيفي المفتوح (OAIS Open Archival Information System-)(CCSDS 2012) وأن يمتلك المســؤولية الإدارية، والســلامة التنظيمية، والاســتدامة المالية، والملائمة التقنية والإجرائية، وأمن النظام، والمســاءلة الإجرائية، وتساعد كل هذه الخصائص معاً على ضمان وبناء الثقة في البيانات ومستودعها.

وتُعــد البنية التحتية الســيبرانية أكثر مــن مجرد مجموعة من المحفوظات الأرشــيفية الموثوقــة فهي تُمكّن من تبادل البيانات والتشــارك في موارد البنيــة التحتية باهظة الثمن. وتتألف البنية التحتية الســيبرانية من مجموعة متكاملة من المعدات المتخصصة والتقنيات





الشبكية (تشمل المراصد والمرافق البحثية التجريبية، وشبكات الاستشعار، وما إلى ذلك)، والخدمات والأدوات لبرمجيات التشغيل المتبادل والبرمجيات الوسيطة، والخدمات المشتركة (المتعلقة بإدارة البيانات ومعالجتها)، وأدوات التحليل مثل التمثيل البصري للبيانات (على سبيل المثال، تطبيقات الرسوم البيانية ورسم الخرائط)، والسياسات المشتركة (فيما يتعلق بحق الوصول للبيانات واستخدامها وحقوق الملكية الفكرية، وغيرها) التي تمكّن من نشر البيانات وتوزيعها، وربطها بعضها مع بعض وتحليلها (هيئة البنية التحتية السيبرانية Cyberinfrastructure Council 2007). كما تسعى إلى تعزيز تنمية المجتمع الثقافي الذي يدعم التعاون المباشر بين الأفراد متجاوزاً الحدود الجغرافية والمؤسسية (Cyberinfrastructure Council 2007). وبينما تستخدم أحياناً لوصف البنية التحتية التي تمكّن عمل مستودع بيانات موحد، فإنها تستخدم هنا للدلالة على البنية التحتية للبيانات التي تتشارك في مواصفات تقنية مشتركة متعلقة بالنماذج، والمعايير، والبروتوكولات. وبعبارة أخرى، هناك قواعد صارمة متعلقة بتطبيق المعايير على البيانات والالتزام بها داخل البنية التحتية. وتشمل مثل هذه البنى التحتية السيبرانية تلك التي تنفذها الوكالات الإحصائية الوطنية والبنى التحتية المكانية الوطنية للبيانات، إضافة إلى البنى التحتية العلمية واسعة النطاق التي تتعلق بأبحاث الجينات، وعلم الفلك، والأرصاد الجوية وغيرها، والتي تتطلب تخزين كافة البيانات ومشاركتها مع الامتثال لمعايير محددة من أجل تحقيق أقصى قدر من قابلية التشغيل البيني للبيانات وضمان جودة البيانات، والإخلاص والنزاهة التي تعزز الثقة. وهذه الأخيرة لها أهمية خاصة في مجال العلوم، والسياسات التي تقوم على أساسها مثل تتبّع تغيرات المناخ وتنبؤاته، حيث تعد سلامة البيانات أمراً بالغ الأهمية في المناقشات الجارية حالياً.

وإضافة إلى بوابات البيانات، والمستودعات، والبنى التحتية، تم استحداث عدد من المنظمات والهياكل المؤسسية التي تقدم المشورة بشأن أفضل الممارسات في مجال الحفظ الرقمي، وإتاحة ومشاركة البيانات، والتحقق من صحة العمل والمعايير المتعلقة بالأنواع المختلفة من بنوك البيانات، والدفع باتجاه تنميتها وتطويرها، وتنسيق الجهود المختلفة بشأن ذلك (انظر الجدول 3-2). وتؤدي هذه المنظمات والهياكل المؤسسية عملها من خلال أنشطة عدة مثل تقديم البرامج التدريبية والتعليمية، واستضافة الندوات والمؤتمرات، وإنتاج

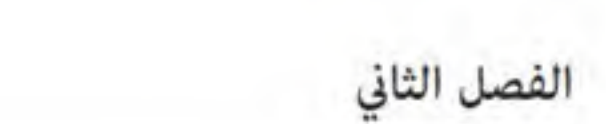


الكتيبات والأدلة، وتوفير التمويل اللازم لمبادرات محددة. وتكتسب هذه المنظمات أهمية نظراً لإسهاماتها في خلق مجتمعات الممارسة بين مختصي البيانات، وتسهيل المفاوضات بين الهيئات وتوسيع نطاق الجهود المؤسسية، ووضع مبادئ توجيهية ومعايير مشتركة. ويركز بقية هذا القسم على مستودعات البيانات البحثية والبنى التحتية، ويفصّل في الأسس المنطقية لها والتحديات التي تواجهها، كما ستتم مناقشة البنى التحتية الخاصة والمتعلقة بشركات البيانات التجارية في وقت لاحق في هذا الفصل.

جدول رقم (٢-٢)

أنواع وأمثلة البنى التحتية للبيانات

النوع	الأمثلة
حاوية البيانات Data Holding	يُطلق على مجموعات غير رسمية من ملفات البيانات التي تكون على جهاز حاسب شخصي، ويسمى نظام النسخ الاحتياطي في بعض الأوقات.
محفوظ البيانات (أرشيف البيانات) Data Archive	أرشيف لوني لبيانات الصور (LONI Image Data Archive-IDA) هو أرشيف آمن لصور الجهاز العصبي والبيانات السريرية الطبية ذات الصلة بالمشاريع البحثية في جميع أنحاء القارة الأمريكية الشمالية والقارة الأوروبية. أرشيف الموسيقى العالمية الموجود في مكتبة موسيقا لويب (Loeb) في جامعة هارفارد الذي يقدّم تخزيناً آمناً ووصولاً متاحاً لمحتوى الأرشيف والتسجيلات التجارية الصوتية والمرئية للموسيقى من جميع أنحاء العالم.
الكتالوجات، الأدلّة، البوابات، البيانات الوصفية، والمجمّعون Catalogues , Directories , Portals , Metadata , Aggregators	تمكّن خدمة البيانات الوطنية الأسترالية المستخدمين من تصفح أكثر من أربعمائة ألف من مجموعات البيانات وأكثر من ثلاثمائة ألف مشروع تقوم بإنشاء مجموعات البيانات. تتيح المكتبة الرقمية العامة الأمريكية الوصول إلى مليونين وخمسمائة ألف سجل من سجلات البيانات الوصفية من مكتبات، ومحفوظات، ومتاحف مقرها الولايات المتحدة. يوروبيانا هي بوابة إلكترونية ومجمّع للبيانات الوصفية لأكثر من مئة وثلاثين من المكتبات والمتاحف والمكتبات الرقمية للوسائط المتعددة في جميع أنحاء أوروبا. محفوظات بيانات المجلس الأوروبي للعلوم الاجتماعية (CESSDA) هي بوابة إلكترونية ومجمّع للبيانات الوصفية منذ ما يزيد عن ثلاث وعشرين سنة لمستودعات البيانات الأوروبية الإحصائية للعلوم الاجتماعية التي تضم أكثر من ثمانية وعشرين ألف من مجموعات البيانات.

مستودعات بيانات الموقع الواحد Single-site Repositories	خدمات أرشفة وشبكات البيانات (DANS) هو الأرشيف الوطني الهولندي لبيانات البحوث الرقمية. إضافة للأرشفة، وتقدّم المبادئ التوجيهية للتدريب والاعتماد. أرشيف بيانات المملكة المتحدة (UKDA) هو مستودع بيانات الموقع الواحد لبيانات بحوث العلوم الاجتماعية والإنسانية الممولة من قبل مجلس البحوث الاقتصادية والاجتماعية (ESRC)، ولجنة نظم المعلومات المشتركة (JISC)، وجامعة إسيكس (Essex).
مستودعات بيانات متعددة المواقع Multi-site Repositories	المستودع الرقمي الآيرلندي (DRI) هو مستودع رقمي موثوق للبيانات المعاصرة والتاريخية، الاجتماعية والثقافية، وتقوم عليه المؤسسات الأيرلندية، ويوفّر إتاحة الوصول، والحفاظ على البيانات للجهات المعنية ذات المصلحة واكتشافها. أوبونير (Openaire) هو مستودع رقمي متعدد المواقع للحصول على البيانات العلمية المفتوحة في منطقة البحث الأوروبية المكوّنة من واحد وأربعين منظمة وثلاثة وثلاثين بلداً أوروبياً. المكتبة الرقمية لتكساس (TDL) هي مجموعة من مؤسسات التعليم العالي التي توفّر منصة وخدمات مشتركة لتوفير الوصول المفتوح للبيانات والحفاظ عليها.
البنى التحتية السيبرانية Cyber- Infrasrtuctures	توفّر البنية التحتية للبيانات الجغرافية المكانية الكندية (CGDI) الوصول الموحّد، والمتسق، والدقيق، والقابل للتشغيل المتبادل للبيانات المكانية عبر القطاعات والنطاقات الكندية.

المصدر : مقتبس بتصرف من (2013).O'Carroll et a!.

جدول (٣-٢)

مجموعة منتقاة من مؤسسات تقديم المشورة، والضغط لكسب التأييد، والتنسيق للمحافظة على البيانات وإتاحتها للوصول والمشاركة وإعادة الاستخدام في العلوم الاجتماعية والإنسانية

المؤسسة بالعربية	الموقع الالكتروني للمؤسسة	المؤسسة بالإنجليزية
اتحاد جميع الأكاديميات الأوروبية	www.allea.org	The Federation of All European (Academies (ALLEA
آريادي‍و	www.ariadne-eu.org	Ariadue
المحفوظات الصوتية البريطانية والايرلندية	www.bisa-web.org	British and Irish Sound Archives ((BISA
الاتحاد الأوروبي لمحفوظات البيانات للعلوم الاجتماعية	www.cessda.org	Consortium of European Social Science Data Archives
التحالف من أجل معلومات مترابطة	www.cni.org	(Coalition for Networked Information (CNI
داتابيب	www.databib.org	Databib
داتاون	dataone.org	DataONE
موارد اللغة المشتركة والبنية التحتية التقنية	www.clarin.eu	Common Language Resources and (Technology Infrastructure (CLARIN
لجنة البيانات للعلوم والتقنية للمجلس الدولي للعلوم	www.codata.org	Committee on Data for Science and Technology of the International Council for Science
الخدمات الرقمية للبنية التحتية للعلوم الاجتماعية والإنسانية	www.dasish.eu	Digital Services Infrastructure for Social (Sciences and Humanities (DASISH
مركز كرأيشن الرقمي	www.dcc.ac.uk	(Digital Curation Centre (DCC
اتحاد المكتبة الرقمية	www.diglib.org	(Digital Library Federation (DLE
تحالف الحفظ الرقمي	www.dpconline.org	(Digital Preservation Coalition (DPC
الشراكة للإنتاج الرقمي	www. digitalproductionpartnership. co.uk	(Digital Production Partnership (DPP



الاتحاد الدولي لمحفوظات التلفزيون	www.fiatifta.org	International Federation of Television (Archives (FIAT/IFTA
الرابطة الدولية لخدمات المعلومات والتكنولوجيا للعلوم الاجتماعية	www.iassistdata.org	International Association for Social Science Information Services and (Technology (IASSIST
شبكة أكاروس	www.icarusweb.arhiv.hr	IcarusNet
الرابطة الدولية للمحفوظات الصوتية والسمعية - البصرية	www.iasa-web.org	International Association of Sound and Audiovisual Archives
اتحاد الجامعات للبحوث السياسية والاجتماعية	www.icpsr.umich.eda/icpsrweb/ICPSR	Inter-university Consortium for (Political and Social Research (ICPSR
شبكة الخبرة في التخزين على المدى الطويل للموارد الرقمية	www.langzeitarchivierung.de//Subsites/nestor/DE/Home	Network of Expertise in Long-term (Storage of Digital Resources (NESTOR
مؤسسة الكواكب المفتوحة	www.openplanetsfoundation.org	Open Planets Foundation
مركز بريستو	www.PrestoCentre.eu	PrestoCentre
ائتلاف بيانات البحوث	www.rd-alliance.org	(Research Data Alliance (RDA
مبادرة ترميز النصوص	www.tei-c.org	(Text Encoding Initiative (TEI

المصدر : مقتبس بتصرف من O'Carroll et al. (2013).

الأسس المنطقية للبنى التحتية للبيانات البحثية:

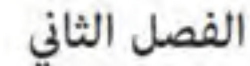
بالنظر إلى أن بناء البنى التحتية للبيانات هي ممارسة ليست باليسيرة، ومكلفة من حيث الموارد والوقت الذي يستلزمه إنشاؤها والمحافظة عليها بشكل مستمر، فينبغي أن تكون هناك أسباب معتبرة لإجراء مثل هذا المسعى والمضي فيه. ويستعرض الجدول (٤-٢) قائمة لأكثر من ثلاثين من هذه الأسباب مقسمة على ثلاثة أبعاد – مباشرة / وغير مباشرة، وفي المدى القريب / وعلى المدى البعيد، والعامة / والخاصة بحسب تعريف Beagrie et al. (2010).



ويُمكن أن تترجم هذه الأسباب تقريبياً إلى المكاسب العلمية والمالية، والتأثير التراكمي للفوائد، والمستفيدين من هذه البنى التحتية.

وتتمحور المناقشات والحجج العلمية لتخزين، ومشاركة، وتوسيع البيانات داخل البنى التحتية للبيانات حول وعود من الاكتشـافات والابتـكارات الجديدة من خلال الجمع بين مجموعات البيانات والتوظيف الجماعي للعقول المشـاركة. فمجموعات البيانات الفردية ذات قيمـة في حـدّ ذاتها، ولكن عنـد ضمّها جنباً إلى جنب مـع مجموعات بيانات أخرى أو تفحصّهـا بطـرق وأسـاليب جديدة، فإن رؤى وأفـكار جديدة مـن المحتمل أن تظهر وأسـئلة جديدة من الممكن أن تتم الإجابة عليها (Borgman 2007)، فبسـبب الطبيعة التراكمية، من خلال الجمع بين مجموعات البيانات، يُمكن الزعم بتسارع وتيرة بناء المعرفة (Lauriault et al. 2007). وعـلاوة عـلى ذلك، يصبح من الممكن من خلال الحفاظ على البيانات تتبع الاتجاهات والأنماط مع مرور الوقت، وكلما ازدادت فترة التسجيل للبيانات، زادت معها القدرة على بناء النماذج وأنظمة المحاكاة ومن ثم زيادة الثقة في الاستنتاجات التـي يتمّ التوصل إليهـا (Lauriault et al. 2007). لذا، فإنه ومـع مرور الوقت، تزداد القيمة التراكمية لمسـتودعات البيانات والبنى التحتية بحيـث تصبح البيانات متاحة على نطاق واسع وبشكل أكثر يسراً، وذلك من حيث النطاق والصلاحية الزمانية.

ومن المرجح أن تعمل هذه المشاركة الاستراتيجية إلى بث روح التعاون بين الباحثين والفرق في التخصصـات الجديدة، وتعزيز المهارات بصورة أكبر من خلال إمكانية الوصول إلى أنواع جديدة من البيانات (Borgman 2007). وعلاوة على ذلك، فإن تبادل المعلومات واعتماد معايير وبروتوكولات وسياسات البنية التحتية من شأنه أن يرفع من مستوى جودة البيانات وأن يُفسـح المجال لأطراف أخرى لإنتاج بيانات مُشتقة والتحقق من البيانات المتاحة، ومن ثم زيادة سـلامة البيانـات (Lauriault et al. 2007). كما تعمل البنى التحتية أيضاً على إتاحة بيانات عالية الجودة في التعليم وتحسين المصادر التربوية والتعليمية لتدريب الجيل القـادم من الباحثين. كما أن البنى التحتية للبيانات هي الوسـيلة الوحيدة لإدارة البيانات والمشـاريع بكفاءة وفعالية في الحالات التي تعتمد فيها المشـاريع العلمية بشكل كبير على أدوات تقنيـة المعلومات والاتصـالات الباهظة الثمن، والتي ينتـج عنها كميات هائلة من

البيانات، وكذلك في المشـاريع التي تضم فرقاً بأعداد كبيرة من الباحثين المنتشرين في جميع أنحـاء العالم لتحليل هذه البيانات، لذا فإن البنى التحتية للبيانات تُمثل الوسـيلة الوحيدة الفاعلة والفعالة لإدارة مثل هذه النوعية من البيانات والمشاريع (Borgman 2007).

وتتمحـور الفوائد المالية للبنى التحتية للبيانات حول الموازيين الاقتصادية التي تنشـأ من خلال تقاسم الموارد ومشاركتها وتجنب تكرارها، والآثار المترتبة على إعادة الاستخدام للبيانـات المكلفـة، والثروة الناتجة من خلال الاكتشـافات الجديـدة، وإنتاج مجتمعات أكـثر كفـاءة. فالبحث عملية مكلفة في العـادة، حيث تنفـق وكالات التمويل المختلفة بشـكل جماعي مليارات الدولارات سنوياً لتمويل النشاط البحثي، وينطوي جزء كبير من هذا النشـاط عـلى ازدواجية غير ضرورية في جهود إنتاج البيانـات، ومعالجتها، وإدارتها، وتحليلها. وبدلاً من إنشاء عدد كبير من المحفوظات الأرشيفية الخاصة، فإنه من المنطقي إنشاء عدد أقل من البنى التحتية المخصصة التي تقوم بتقديم وتحضير البيانات الأساسية عـلى صورة موحّدة، وتُنتـج فعالية كبيرة في الجهد، وتُمكن كذلك وصولاً أوسـع للبيانات للأفراد الباحثين والمؤسسـات كـون تكاليف الدخول إلى حقل أو مجـال بحثي في العادة باهظة (Fry et al. 2008). وإضافة إلى ذلك، ستصبح عمليات ونتائج المشاريع البحثية أكثر شـفافية ويُمكـن الحكم على مدى تمثيلها القيمة مقابل المـال. وفضلاً عن الحدّ من الهدر والحفاظ على الثمار الناتجة من هذه الجهود وتقاسمها، فإنها ستعمل على تعظيم العائد من الاستثمار من خلال تمكين أكبر قدر من المكاسب التي يمكن الحصول عليها من البيانات (Lauriault et al. 2007). فعلى سبيل المثال، قد ينتج عن إتاحة البيانات على نطاق واسع صناعات جديدة قائمة على المعرفة وتكون شركات منبثقة ومنفصلة من خلال تسـويق البحوث الممولة من القطاع العام (ائتلاف العلوم Science Coalition 2013)، كما يمكن للبيانات ضمن البنى التحتية للبيانات أن تؤدي إلى طرق أكثر فعالية في الحوكمة والإدارة المجتمعية. على سـبيل المثال، توفر العديد من البنـى التحتية للبيانات المكانية الوطنية، إضافـة إلى دعمها للبحوث المكانية، أدوات فعالة لـلإدارة والحكومة الإقليمية، وتمكّن من الاستعداد للطوارئ والاستجابة للكوارث بصورة أكثر فعالية، وتُسهيل التخطيط بين الدول.

وبحسب ما تم إيضاحه من خلال النقاش حتى الآن، فإن الفوائد المحتملة المكتسبة من البنى التحتية للبيانات هي فوائد ممتدة وغير محصورة بالقطاع العام أو الخاص (Beagrie et al. 2010). وقد ظهر جلياً في الواقع أن استراتيجية مشاركة البيانات وتقاسمها تعود بالفائدة على جميع الأطراف – الباحثين، المؤسسات، الممولين، والمؤسسات العامة والخاصة. فعلى مدى العقدين الماضيين، وبالنظر إلى المكاسب المتوقعة من مشاركة البيانات وتبادلها، عملت الوكالات البحثية للحكومات الوطنية والهيئات الإقليمية مثل الاتحاد الأوروبي، جنباً إلى جنب مع المنظمات الخيرية، على الاستثمار بنطاق واسع في تمويل مجموعة كبيرة من المبادرات المختلفة. ومع ذلك، وكما هو مبيّن أدناه، فهناك مقاومة لمثل هذه المشاريع، وعلى الأخص من قبل الباحثين أنفسهم، ومن الواضح أنه على الرغم من أن هناك فوائد لا شكّ فيها، فهناك أيضاً تكاليف مرتبطة بإنشاء البنى التحتية للبيانات.

تحديات بناء البنى التحتية للبيانات:

على الرغم من الفوائد المتوقعة والمُدركة من البنى التحتية للبيانات، إلا أنها لم تلق ترحيباً عالمياً، كما أنها أثبتت صعوبة بنائها وتنفيذها، فهناك مجموعتين رئيسيتين من القضايا الفاعلة - التقنية والاجتماعية/ السياسية - التي تعوق تطوير البنى التحتية للبيانات وتنميتها، وستواصل القيام بذلك لبعض الوقت.

جدول رقم (٤-٢)

فوائد مستودعات البيانات والبنى التحتية للبيانات

الفوائد المباشرة	الفوائد غير المباشرة (تجنب التكاليف)
فرص جديدة للبحث.	عدم تكرار إنشاء البيانات / الازدواجية في البيانات
التواصل العلمي / إتاحة الوصول إلى البيانات.	عدم فقدان لفرص البحث المستقبلية.
إعادة تحديد الأهداف وإعادة استخدام البيانات.	انخفاض تكاليف المحافظة على البيانات في المستقبل.
زيادة الإنتاجية البحثية	إعادة توظيف البيانات للمهتمين الجدد.
تحفيز أنواع جديدة من التواصل / التعاون.	إعادة توظيف المنهجيات المستخدمة.
إتاحة البيانات للتعليم ومشاريع الطلاب.	الاستخدام من قبل مهتمين جدد.

نقل المعرفة للصناعة.	حماية العائد على الاستثمار المبكر.
تحسين قاعدة المهارات.	الأدوات والمعايير يحتمل أن تزيد من مستوى جودة البيانات.
زيادة الإنتاجية / النمو الاقتصادي.	تقليل الاستفسارات العشوائية المتعلقة بالبيانات
التثبت من البحث / سلامة البحث	
الوفاء بالالتزامات	
فوائد على المدى القريب	**الفوائد على المدى الطويل**
قيمة للباحث والطلاب الحاليين.	تأمين القيمة للباحثين والطلاب المستقبليين.
عدم فقدان البيانات بتغير الباحثين.	إضافة القيمة مع مرور الوقت ونمو المجموعة وتطورها للحد الأقصى.
توسيع إتاحة الوصول حيث التكاليف الباهظة للباحثين / المؤسسات.	زيادة سرعة البحث والوقت اللازم لإدراك آثاره.
إعادة الاستخدام على المدى القصير للبيانات المنسقة جيداً.	يحفز الأسئلة البحثية الجديدة، وبخاصة المتعلقة بالبيانات المرتبطة والمشتقة.
التخزين الآمن لبيانات البحوث المكثفة.	
تعزيز توافر البيانات للنشر والمشاركة.	
الفوائد الخاصة	**الفوائد العامة**
فوائد للرعاة / الممولين من البحث / المحفوظات.	مدخل للبحوث المستقبلية.
فوائد للباحثين والمؤسسات.	تحفيز البحوث الجديدة.
الوفاء بالالتزامات الخاصة بالمنح.	تحفيز الشركات الجديدة وتوظيف المهارات العالية.
زيادة الوضوح / الاقتباس.	الشفافية في تمويل البحوث.
الاتجار بالبحوث	

المصدر: Beagrie et al. 2010; Fry et al 2008.



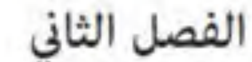
إن إنشـاء بنية تحتية عاملة وقابلة للتشغيل المتبادل للبيانات ليس تحدّياً هيّناً، خاصة إذا كان النظام موثوقاً ويقدّم مسـتويات مختلفة مـن إتاحة الوصول اعتماداً على تعريف المسـتخدم وضمانات فيما يتعلق بمصدر البيانات وسـلامتها، ونزاهتهـا، وصلاحيتها، فذلك يتطلـب تعاوناً مكثّفاً وواسـع النطاق والموائمـة بين منتجي البيانـات ومالكي المحفوظات الأرشيفية فيما يتعلق بالتالي:

- إجراءات إنتاج البيانات.
- نماذج البيانات ومعايير البيانات.
- البيانات الوصفية والتوثيق.
- إجراءات تحضير وتنقية وإدخال البيانات.
- جودة البيانات وضمانها.
- سياسات الحفاظ على البيانات، والنسخ الاحتياطي، والتدقيق عليها.
- البرمجيات والمعدات.
- أمن وحماية المعلومات.
- سياسـات الوصول، ومنح التراخيص، والاسـتخدام، وإعادة الاسـتخدام، والخصوصية، والجوانب الأخلاقية.
- سياسة الملكية، وحقوق الطبع والنشر، وحقوق الملكية الفكرية.
- الترتيبات الإدارية وآليات تنظيم الإدارة والحوكمة.
- تمويل البنية التحتية وخدماتها وإدارتها.

(Borgmann 2007; Data Seal of Approval 2010; O’Carroll and Webb 2012)

فالبنيـة التحتية للبيانات لن تنجح إلا إذا كان لديهـا رؤية طويلة الأجل، وهناك التزام لإدارة المحتـوى نيابة عن المودعين والمسـتخدمين وتحمّل مسـؤوليتها، مـع التطبيق التام لجميع العوامل المذكورة أعلاه بشكل فعال (O’Carroll and Webb 2012).





فالقضايا التي تمت الإشارة إليها ليست مجرد قضايا تقنية، ولكنها أيضاً اجتماعية وسياسية، ذلـك لأنها تتطلب التفاوض والاتفاق عـلى طريقة تصميمها واختيارهـا ومتطلبات إدارتها القائمة بـين العديد من الأطراف - الممولـين، والمصممين، والمبدعـين، والمديرين والمودعين، ومجموعات الضغط والتوجيه، والمسـاهمين وأصحاب المصالح الخاصة الأخرى - في سـياق مختلف التعليمات، واللوائح، والقوانين المتعلقة بإنتاج، وتخزين، ومشـاركة البيانات. ومن ثم، فهي تتشـكّل بواسطة الأجندات الشخصية والشخصيات، والقيم المؤسسية، والالتزامات والوصايـات، والثقافات، والهيـاكل، والعمليات، والتمويل، والقضايـا الحكومية، والتبعيات الإجرائية حيث أنه بمجرد أن يبدأ المشروع في مسار معين، فسيكون من الصعب غالباً توجيهه في مسـار مختلـف كلياً عنــه (Feenberg in Lauriault 2012; Lauriault et al. 2007). كـما يمكن أن يكون تجميع ومواءمة المحفوظات الأرشـيفية المعتمدة على مسـارات تقنية مختلفـة عملية في غاية الصعوبـة، وطويلة الأمد، ومكلفة لأنها تتطلب من تلك المشـاريع إجراء تغيـيرات لتحديث بياناتها المتقادمة لتتوافق مع المواصفـات المتفق عليها في المواءمة (Lauriault 2012). وهذا هو السـبب في اعتماد العديد من المشـاريع الأرشيفية على بنية مستودع البيانات بدلاً من السعي لتصبح بنية تحتية إلكترونية/ سيبرانية.

ويـأتي بعد الصعوبات المتعلقة بإنشـاء هيـاكل تبادل البيانات، إقنـاع أصحاب البيانات أنفســهم بتبادل سلعتهم الثمينة (بياناتهم)، فهناك مبدأ أسـاسي في البحوث الأكاديمية وهو إتاحة الوصول المفتوح؛ بمعنى، أن جميع جوانب إنتاج المعرفة ينبغي أن تكون متاحة بحرية للآخرين لتفحصها ونقدها. وهذا المبدأ لم يُعمل به على النحو الأمثل أبداً، إذ تختفي الأبحاث الأكاديميـة في الغالب خلف حواجز الأجور ويتردد الباحثون في مشـاركة بياناتهم وتبدلها مع الآخرين. وكما أشـار (2007) Borgman، فإن مشـاركة البيانات وتبادلها شائع في عدد قليل من التخصصات مثل علم الفلك (Astronomy)، علم الجينات (Genomics)، وعلم البيانات العمرانية بهيئتها الرقمية (Geomatics) والتي تعتمد على فرق علمية كبيرة وموزعة جغرافيا؛ أما في التخصصات الأخرى فيجري في أحيان نادرة مشاركة البيانات وقد لا تجري على الإطلاق. وقد خلصت Borgman إلى أن «السرّ الصغير القذر» وراء تشجيع مشاركة البيانات، وتبادلها هـو عدم التبادل لكثير من البيانـات في الواقع (Borgman 2012: 1059). وذلك لأن هناك عدداً من المثبطات التي تحول دون تبادل البيانات ومنها:

- عدم وجود مكافآت للقيام بذلك.
- الجهد اللازم لإعداد وأرشفة البيانات.
- نقص الخبرة والموارد والأدوات اللازمة لأرشفة البيانات.
- مخاوف من القدرة على استخراج قيمة من البيانات قبل الغير، وذلك من حيث الأوراق العلمية وبراءات الاختراع، ونظراً للجهد المستثمر في إنتاج البيانات.
- المخاوف بشأن الكيفية التي سيتم بها استخدام البيانات، خاصة إذا كانت تتعلق بأشخاص، أو كيف يمكن أن يساء التعامل معها أو يساء تفسيرها.
- المخاوف بشأن الاستفسارات بخصوص إنتاج البيانات والطلبات التي من شأنها أن تستلزم أعمالاً إضافية.
- مخاوف من القضايا المتعلقة بالكشف عن البيانات ونتائج البحوث التي يمكن تقويضها من خلال تفسيرات بديلة لنفس البيانات.
- قضايا الملكية الفكرية.
- الخوف أنه لن يتم استخدام البيانات، وبالتالي تشكّل الأرشفة جهداً ضائعاً. (Borgman 2007, 2012; Strasser 2013).

فمن الواضح أن ضمان البيانات من خلال أرشفتها لإعادة استخدامها مستقبلاً سوف يتطلب أكثر من مجرد إنشاء مستودعات البيانات والبنى التحتية لها؛ كون ذلك يتطلب تغييراً ثقافياً في الممارسات البحثية. ويتكشف هذا التغيير للعيان باستخدام إستراتيجية العصا والجزرة، فمن ناحية بدأ استخدام الحوافز لتشجيع الباحثين على أيداع وتبادل البيانات، مثل تشجيع الاقتباس والإسناد عند إعادة استخدام البيانات (Borgman 2012)، وتوفير الأموال الكافية لأرشفة البيانات ضمن جوائز المنح. ومن ناحية أخرى، بدأت وكالات الأبحاث في إجبار الباحثين على إيداع وتبادل البيانات، مع مراعاة الجوانب الأخلاقية وحقوق الملكية الفكرية، كشرط لتمويل البحوث.



وسطاء البيانات والأسواق (Data Brokers and Markets):

يمتلك القطاع الخاص دافعاً مختلفاً جداً لإنتاج وإعادة استخدام البيانات عن ذلك الذي عند الأكاديميين، فالبنى التحتية للبيانات البحثية تسعى لتحقيق الصالح العام من خلال إنشاء مشاعات للبيانات لتسهيل وتشجيع تبادل البيانات وإعادة استخدامها. وفي المقابل، يلتقط وسطاء البيانات (يسمون أحياناً مجمعي أو مختزلي أو بائعي البيانات) البيانات ويجمعونها معاً ويعيدون تعبئتها (تحزيمها) في البنى التحتية للبيانات التي يملكها القطاع الخاص لعرضها للإيجار (للاستخدام لمرة واحدة أو لاستخدامها تحت شروط الترخيص) أو إعادة البيع على أسس ربحية. وتشمل المنتجات قوائم العملاء/الزبائن المحتملين الذين يستوفون معايير معينة (بما في ذلك تفاصيل مثل الأسماء، العناوين، أرقام الهاتف، وعناوين البريد الإلكتروني، وكذلك معلومات مثل «الجنس، العمر، العرق، وجود الأطفال، الدخل، قيمة المنزل، ملكية بطاقة الائتمان، الوضع الائتماني، أنماط الشراء، والهوايات»؛ (CIPPIC 2006: ii)، والتدقيق على الخلفية الأمنية، ومجموعة من منتجات البيانات المشتقة التي أضاف لها الوسطاء قيمة من خلال التكامل، ومنتجات تحليل البيانات التي تستخدم في معظم الأحوال لحملات الدعاية والتسويق المصغرة، وتقييم الجدارة الائتمانية والتصنيف الاجتماعي للأفراد، وتوفير خدمات البحث عن المفقودين، وتقديم تحليلات الأعمال المفصّلة (CIPPIC 2006). وفي حالة القوائم، فهي في معظمها مستأجرة وفي كثير من الحالات لا يستلم المستأجر القائمة، حيث يعمل مكتب خدمة كطرف ثالث على إعداد وإرسال البريد بالنيابة عنهم (CIPPIC 2006). وبالمثل، مع تحليلات البيانات، حيث يتلقى العملاء نتائج تحليل البيانات، وليس ذات البيانات الأساسية.

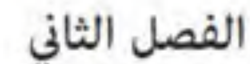
إن عمليات اختزال البيانات وإعادة بيعها، وما يرتبط بها من تحليل للبيانات وخدمات ذات قيمة مضافة، هي صناعة بمليارات الدولارات، مع كميات هائلة من البيانات والمعلومات المستمدة التي تُؤجر، وتُباع، وتُشتري بشكل يومي عبر مجموعة متنوعة من الأسواق - تجارة التجزئة، والمالية، والصحة، والسياحة، والخدمات اللوجستية، والأعمال الذكية، والعقارات، والأمن الخاص، والاقتراع السياسي، وإلى آخره. وتتعلق هذه البيانات بجميع جوانب الحياة اليومية وتشمل الإدارة العامة، والاتصالات، واستهلاك السلع ووسائل الإعلام، والسفر،



والترفيــه، والجريمة، وتفاعلات وســائل التواصــل الاجتماعية، وإلى آخره، فوســطاء البيانات المتخصصين هم متواجدون منذ فترة طويلة، يجمعون البيانات من اشــتراكات وسائل الإعلام (على سبيل المثال، الصحف، والمجلات)، وتجار التجزئة عبر طلبات البريد، واستطلاعات الرأي، والدراسات الاســتقصائية، ووكالات السفر، والمؤتمرات، والمســابقات، وتسجيلات وضمانات المنتجات، وشركات معالجة الدفع والســداد، والســجلات الحكوميــة، وإلى آخره (CIPPIC 2006). وفي الســنوات الأخيرة، وخاصة منذ ظهور الإنترنت، توســعت هذه الصناعة بشكل سريع، مع التنوع في البيانات التي تم التقاطها واختزالها، والمنتجات والخدمات التي تقدمها، وغالباً ما تجمع معا على حدّ سواء البيانات الصغيرة والكبيرة.

فبيع البيانات للوسطاء أصبح يمثل تدفقاً مهما للدخل للعديد من الشركات، فعلى سبيل المثــال، يبيع تجار التجزئة، في كثير من الأحيان، البيانــات المتعلقة بالعمليات مثل تفاصيل بطاقة الائتمان، ومشــتريات العملاء وبرامج الولاء الخاصة بالمتجر، وإدارة علاقات العملاء، ومعلومات الاشــتراك. كــما تبيع المواقع الإلكترونيــة بيانات النقر (الاختيــارات) المتعلقة بكيفيــة تصفح شــخص ما لموقع على شــبكة الإنترنت والوقت الــذي يقضيه على مختلف الصفحــات. وبالمثل تســتخلص شركات وســائط الإعلام، مثل محطــات الصحف والإذاعة والتلفزيون، البيانات الواردة في محتواها (على سبيل المثال، الأخبار، والإعلانات). وتستخلص شركات وســائط التواصل الاجتماعــي البيانات الوصفيــة وبيانات المحتوى لمســتخدميها، وتســتخدم هذه البيانات لتحديد المواصفات الأساســية للمســتخدمين ومنتجات البيانات الخاصة بهم، أو بيع البيانات لوســطاء البيانات. على سبيل المثال، تستخدم شركة فيسبوك الملفــات التعريفية للمليارات من مســتخدميها، وشــبكاتهم الاجتماعيــة، والمحتوى الذي يقومون بتحميله (المفضــلات، والتعليقات، والصور، والفيديوهــات، وما إلى ذلك) لتمكين مجموعــة من المنتجات الإعلانية مثل الإعلانات الموجهــة للجماهير المتماثلة (Lookalike Audiences)، والجماهير المخصصة المــدارة (Managed Custom Audiences)، وفئات الشريك (Partner Categories) وذلك بالشراكة مع وســطاء البيانات الكبيرة والمســوقين مثل داتا لوجيكس (Datalogix)، وأبســيلون (Epsilon)، وأكسيوم (Acxiom)، وبلو كاي (BlueKai) مــن أجل دمج بيانات الشراء الخاصة بهم التي لا تخص فيســبوك مع البيانات السلوكية التي يوفرها الفيسبوك (Edwards 2013).





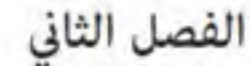
فبطرق مختلفة، يتخلى الأفراد عن البيانات الشخصية الخاصة بهم، عن قصد أوعن غير قصد، وبدرجات متفاوتة: بصفة مشترين، أو مشتركين، أو مسجلين، أو أعضاء، أو حاملي بطاقات، أو مانحين، أو متسابقين، أو مشاركين في الاستطلاعات، أو حتى مجرد مستفسرين (CIPPIC 2006: ii). وعلاوة على ذلك، وحيث إن إنشاء، وإدارة، وتحليل البيانات هي مهمة متخصصة، تستعين العديد من الشركات بجهات خارجية من شركات معالجة البيانات وتحليلاتها للوفاء بمتطلبات البيانات الخاصة بها. وتستطيع هذه الشركات، من خلال تقديم هذا النوع من خدمات البيانات لمختلف العملاء، بناء مجموعات بيانات واسعة النطاق، والتي يمكن الجمع بينها وحزمها لاستخدامها في إنتاج بيانات مشتقة جديدة، والتي بدورها توفر المزيد من الأفكار والرؤى مقارنة بتلك التي يمكن تحصيلها من مصدر واحد فقط للبيانات، وإضافة إلى هذه البيانات التي يحصل عليها وسطاء البيانات من القطاع الخاص، يعمل هؤلاء الوسطاء كذلك على الحصول على مجاميع بيانات القطاع العام سواء التي تتعلق بالأفراد أو تلك التي تخص مجموعة منهم كفئة معينة وأماكنهم مثل بيانات سجلات الملكية والتعداد السكاني، كما يقومون باستئجار/ أو شراء القوائم المتعلقة بالأفراد من الجمعيات الخيرية والمنظمات غير الحكومية.

ومن خلال تجميع البيانات من مصادر متنوعة، يُنشئ وسطاء البيانات بنية تحتية من البيانات الضخمة والمترابطة، فعلى سبيل المثال، تشتهر شركة إبسيلون (Epsilon) بامتلاك بيانات عن ثلاثمائة مليون بطاقة ولاء من أعضائها في جميع أنحاء العالم، إضافة إلى بنك معلومات يضمّ بيانات ذات صلة بمئتين وخمسين مليون مستهلك في الولايات المتحدة وحدها (Edwards 2013). كما تشتهر شركة أكسيوم (Acxiom) بإنشائها بنكاً للمعلومات يتعلق بخمسمائة مليون مستهلك فعّال في جميع أنحاء العالم (تقريباً مئة وتسعون مليون شخص، ومئة وستة وعشرون مليون أسرة في الولايات المتحدة)، مع نحو ألف وخمسمائة معلومة عن كل شخص، وتنفذ خوادمها أكثر من خمسين تريليون عملية في السنة، وتجاوز إجمالي تدوير رأس المال للشركة المليار دولار (Singer 2012a). كما تدير أيضاً قواعد بيانات منفصلة، أو تعمل مع سبعة وأربعين شركة من كبرى الشركات المئة الأكبر في العالم (Singer 2012a). وتزعم شركة داتا لوجيكس (Datalogix) بتخزينها بيانات متعلقة بمشتريات لم تتم عن طريق الإنترنت تبلغ قيمتها أكثر من تريليون دولار (Edwards 2013).



وتشـمل شركات الوساطة والتحليل للبيانات شركات أخرى مثل شركة ألايانس داتا سيستمز (Alliance Data Systems)، وإي بيرو (eBureau)، وتشـويس بوينت (ChoicePoint)، وكورلوجيـك (Corelogic)، وإيكيوفاكس (Equifax)، وإكسـبيريان (Experian)، آي دي أنالاتكس (ID Analytics)، وإنفوجروب (Infogroup)، وإنوفيس (Innovis)، وإنتليوس (Intelius)، وريكورديد فيوتشر (Recorded Future)، وسيسنت (Seisint)، وترانز يونيون (TransUnion). وتتجـه كل شركة من هذه الـشركات إلى التخصص في أنواع مختلفة من البيانـات ومنتجـات البيانات والخدمات، فعلى سـبيل المثال، تقيّم شركـة إي بيرو العملاء المحتملـين بالنيابة عن شركات بطاقات الائتمان، والمقرضين، وشركات التأمين، والمؤسسـات التعليمية، كما تقدّم شركة إنتليوس خدمات البحث عن الأشـخاص والتحري عن خلفياتهم (Singer 2012a).

وبشـكل عام، فإن مبتغى وسـطاء البيانات وشركات التحليل هو مجموعة واسـعة من البيانـات (الصغيرة والكبيرة)، والمتعلقة بأكبر شريحة ممكنة من السـكان، وأن تكون هذه البيانات مترابطة بشـكل كبير ويمكن تمييزها (تحوي تعريف فريد) بحسب طبيعتها، وكلما اسـتطاع وسـيط البيانات تصدير ودمج المزيد من البيانات، كان عمل منتجاته أكثر نجاحاً وعلى النحو الأمثل، واكتسـب ميزة تنافسية عن منافسيه الآخرين. ويُمكن من خلال جمع البيانات معا وهيكلتها بشـكل مناسب لوسـطاء البيانات اشتقاق بيانات جديدة، وملفات تعريفيـة عن الأفراد والمناطق، والقيـام بإجراء النمذجة التنبؤية لتوقع ما قد يفعله الأفراد في ظل ظروف مختلفة وفي أماكن مختلفة. وهذا من شأنه أن يُمكن من تحديد المستهدفين والمخاطر بشـكل أكثر فاعلية بين مختلف المستهلكين وأن يُعطي مؤشراً على مدى التصرف المحتمل لهؤلاء المسـتهدفين والدفع بهم لاختيار اسـتجابة معينة (على سبيل المثال، اختيار وشراء عنصر معين)، فشركة أكسـيوم تسـعى، على سبيل المثال، لربط البيانات غير المتوافرة على الإنترنـت (Offline Data)، والبيانات المتوافرة على الإنترنت (Online Data)، وحتى البيانـات المتنقلـة (Mobile Data) مـن أجل تكوين نظرة شـاملة متعـددة الزوايا عن المسـتهلكين، حيث تستخدم هذه البيانات لإنشـاء ملفات تعريفية مفصلة ونماذج تنبؤية متماسـكة (Singer 2012a). فهذه المعلومات والنـماذج مفيدة جداً لشركات الشراء لأنها تمكنهم من تركيز جهودهم المتعلقة بالتسويق والمبيعات، وتزيد من فرص نجاح عملياتهم،

في ذات الوقت الذي تمكنهم من الحدّ من النفقات من حيث الهدر والخسارة التي قد تنتج عن استثمارات محفوفة بالمخاطر، ومن ثم تسعى هذه الشركات لتصبح أكثر فعالية وكفاءة في عملياتها باستخدامها لمثل هذه المنتجات.

ومن المثير للاهتمام، أنه بالرغم من امتلاك وسطاء البيانات وشركات التحليل لكميات كبيرة ومتنوعة من البيانات الشخصية التي تُستخدم منتجاتها لتصنيف واستهداف الأفراد والأسر اجتماعياً، فقد كان هناك نقص ملحوظ في الدفع بالاهتمام النقدي لعمليات هذه الشركات. ففي الواقع، هنالك ندرة في التحليل الأكاديمي والإعلامي عن هذه الشركات والآثار المترتبة على عملها ومنتجاتها، ويعود السبب في ذلك جزئياً إلى أن هذه الصناعة بعيدة عن الأضواء وسرية نسبياً، ولا ترغب في توجيه انتباه الجمهور إليها، وزعزعة ثقة الجمهور في أصولها وأنشطتها، والتي قد تؤدي إلى حملات عامة للشفافية، والمساءلة، والتنظيم. فوسطاء البيانات بشكل عام غير منظمين إلى حدّ كبير في الوقت الحاضر، كما أنهم غير ملزمين قانونياً بتزويد الأفراد بإمكانية الوصول إلى البيانات المحفوظة عنهم، ولا هم ملزمون بتصحيح الأخطاء المتعلقة بهؤلاء الأفراد (Singer 2012b). وإضافة إلى أن بيانات ومنتجات هذه الشركات يُمكن أن يكون لها تأثير عميق في الخدمات والفرص المقدمة للأفراد، مثل ما إذا كان سيتم تقديم عرض لفرصة عمل، أو قبول تقديم طلب قرض، أو إصدار بوليصة تأمين، أو الموافقة على الاستئجار، وتكلفة السلع والخدمات المقدّرة على أساس المخاطر المتوقعة والقيمة بالنسبة لشركة ما (Lyon 2002).

إن قلق البعض، بمن فيهم Edith Ramirez (2013) رئيسة لجنة التجارة الاتحادية (Federal Trade Commission) في الولايات المتحدة، يتمثّل في ممارسة هذه الشركات شكلاً من أشكال حتمية البيانات (Data Determinism) التي لا يجري فيها فقط تصنيف الأفراد والحكم عليهم على أساس ما فعلوه، ولكن على التنبؤ بما قد يفعلونه في المستقبل باستخدام خوارزميات هي أبعد ما تكون عن الصحة والكمال، بحيث قد تحمل في ثناياها تحيزات تتعلق بالعرق، والإثنية، والجنس، والحياة الجنسية. وحتى الآن، فهم صناديق سوداء تفتقر إلى الرقابة المجدية والإجراءات المُعالجة، وتستخدم بيانات منخفضة الجودة على الأغلب، ومن ثم تكون عرضة للخطأ (انظر أيضاً للفصل العاشر)، وعلاوة على ذلك



فــإن هذه الشركات تســتخدم البيانات لغير الأغــراض التي تمّ إنشــاؤها لأجلها، وبالتأكيد ليست لأغراض أولئك الذين تشير البيانات إليهم وتمثل ما قد يرغبون فيه، فالذين يتقدمون ببياناتهــم للحصــول على بطاقــة ولاء المتجر لا يتوقعــون بالضرورة أن يتم بيــع بياناتهم واستخدامها لأغراض تتجاوز علاقتهم مع المتجر، ولا يتوقعون استخدامها بطرق قد تعرضهم للعقاب والمحاسبة بعد ذلك (Ramirez 2013). وبينما يرحب بعض المستهلكين بالعروض الشــخصية من المتاجر أو الإعلانات الموجّهة على المواقــع الإلكترونية، يعدها البعض الآخر تطفّلاً واســتغلالية (Singer 2012a). وعلاوة على ذلك، فــإن كنز البيانات التي قد يكون لها قيمة في المستقبل كإجراء للمضاربة السوقية، هو انتهاك لقوانين الحدّ من البيانات التي تنصّ على وجوب الإبقاء على البيانات محددة القيمة فقط (Raley 2013). ويُعد وســطاء البيانات هدفاً رئيســياً للمجرمين العازمين على الاحتيــال وسرقة الهويات، وذلك بالنظر إلى حجم السجلات الشخصية الحساسة مثل الأســماء، والعناوين، وأرقام الهوية (مثل الضمان الاجتماعي، وجواز الســفر، ورخصة القيادة)، وأرقام بطاقات الائتمان المتوافرة لديهم. وفي الواقــع، فقد تابعت لجنة التجارة الاتحادية (Federal Trade Commission- FTC) أكثر من أربعين قضية لخروقات أمنية خطيرة للبيانات، أغلبها ضد وســطاء البيانات، لفشلهم في توفير ضمانات أمنية كافية على البيانات (Ramirez 2013).

وفي الآونة الأخيرة، وفي أعقاب الخروقات الأمنية والكشــف عن صفقات بيع مشــبوهة واســتخدام لبيانات حساســة (Dwoskin 2013)، جنباً إلى جنب مع الكشــف عن برنامج بريزم التجسسي لوكالة الأمن القومي (NSA PRISM) ووصول الحكومة إلى سجلات تجارية (Greenwald and MacAskill 2013)، إضافة للأخبار عن فضيحة قرصنة الهاتف العالمي واســتجواب ليفينســون (Leveson) في المملكة المتحدة (http://www.levesoninquiry.org.uk)، فقد انصبّ المزيد من الاهتمام على وسطاء البيانات وممارساتهم التجارية. ففي أواخر عام 2012، اســتدعت لجنة التجارة الاتحادية تســعة من وسطاء البيانات لاكتشاف المزيد حول ماهية البيانات والمعلومات المســتمدة منها التي يقومون بإنتاجها وجمعها عن النــاس وكيفية توظيفهم هذه البيانــات وطرق بيعها، وأصدرت تقريــراً يدعو للخصوصية عنــد التصميم (انظر الفصل العاشر)، وخيارات مبســطة، وقدر أكبر من الشــفافية (لجنة التجــارة الاتحاديــة Federal Trade Commission 2012). وعــلاوة على ذلك، فقد تمّ

تطبيق الضغط السياسي لاعتماد أنظمة أشدّ رقابة وأكثر شفافية. أما في أوروبا، فيسعى قانون الاتحاد الأوروبي الجديد لحماية البيانات لتنظيم كيفية استخدام البيانات، ومنح المواطنين الحق في الوصول إلى البيانات والتنازع عليها والحق في تناسيها وحذفها (انظر الفصل العاشر). وسيعتمد مدى نجاح مثل هذه الأعمال على إعادة تشكيل أعمال وسطاء البيانات والذي سيحدثنا عنه الزمن لاحقاً، وفي غضون ذلك، هناك العديد من الأسئلة المتعلقة بوسطاء البيانات والأسواق وآثارها على المواطنين والمستهلكين التي تحتاج إلى استكشاف للإجابة عنها (انظر الفصل السادس).

الخلاصة:

على الرغم من الدعوات المنمقة لمؤيدي البيانات الكبيرة وأولويات التمويل للوكالات البحثية، فإن البيانات الصغيرة ستستمر بالبقاء باعتبارها عنصراً حيوياً في فضاء البحث العلمي، فدراسات البيانات الصغيرة قد تكون محدودة في الحجم، ومحصورة من حيث عيناتها وتنوعها، وثابتة، وغير مرنة بمجرد تنفيذها، ولكنها خدمت العلوم بشكل جيد حتى وقتنا الحاضر لأنها مصممة خصيصاً للرد على أسئلة مركزة بإحكام. ويُتيح إنتاج البيانات الصغيرة، جنباً إلى جنب، مع أدوات التحليل الجيدة، للباحثين التنقيب الدقيق وبشكل فعّال في مستويات محكمة من البيانات العالية الجودة من أجل فهم العالم. وفي المقابل، وكما نوقش في الفصل الثامن، تسعى العديد من دراسات البيانات الكبيرة إلى الإجابة عن الأسئلة التي يمكن الإجابة عنها من البيانات المنتجة؛ فكما قيل: الذيل الآن هو الذي يهزّ الكلب.

ومع ذلك، فمن الواضح أن رؤى وقيمة أكبر بكثير يمكن الحصول عليها من البيانات الصغيرة من خلال توسيع نطاقها في البنى التحتية للبيانات والاحتفاظ بها للأجيال القادمة. وفي حين أن البيانات ولفترات طويلة قد جُمعت معاً، وجرى تخزينها وأرشفتها، إلا أن تطور التقنيات الرقمية قد غيّر بشكل جذري قدرتنا على إدارة، وتنظيم، ومعالجة، وتحليل، ومشاركة، وإعادة استخدام البيانات، وخاصة تلك البيانات التي ولدت رقمية في الأصل، فقد كانت البيانات الصغيرة فيما سبق نادرة ومقيّدة، ولكنها غدت متوافرة بشكل أكبر وأكثر انفتاحاً (انظر الفصل الثالث). وبتوسيع نطاق هذه البيانات يمكن تحقيق جدوى اقتصادية وذلك فيما يتعلق بخدمات البيانات، وأدوات التحليل والموارد التربوية، وتوظيف

البيانات في اســتخدامات جديدة، وإنتاج اكتشافات جديدة. ونتيجة لذلك، يُبذل الآن جهد كبير لإنتاج البنى التحتية للبيانات في جميع المجالات الأكاديمية. وهو الجهد الذي من شأنه أن يُتيح بشكل أكبر بكثير البيانات للمجتمعات المعنية وأن يُغيِّر طبيعة عمل العلوم (انظر الفصل الثامن).

ويتضح المدى الذي يمكن من خلاله لهذا التوســع إنتــاج قيمة من خلال النمو السريع لقطاعي وساطة البيانات والتحليلات المصاحبة لها. حيث يمكن لوسطاء البيانات، من خلال الجمع بين مخازن كبيرة من البيانات الصغيرة التي تحتفظ بها المؤسسات العامة والشركات الخاصــة ودمجها جنبــاً إلى جنب مع مجالات تدفق البيانات الكبــيرة، إنتاج مختلف أنواع الملفات التعريفية التفصيلية للأفراد والمجموعات والتي يمكن استخدامها للاستهداف المركّز للأسواق، وتقييمها، وفرزها، مما يوفر معلومات ذات قيمة عالية للعملاء. كما تكشف هذه الممارســات عن بعض القضايا والمخاطر الناجمة عن التوســع في نطاق البيانات، وبالتحديد اســتخدامها لغير الأغراض التي جُمعت لأجلها، وأمنها، والفــرص الجنائية في حال تم سرقة البيانــات، حيــث لم يُفهم أي من هذه القضايا والمخاطر بشــكل تام حتى الآن، مع ضعف الجانب التنظيمي والإجراءات المضادة.

فهنــاك العشرات من الأســئلة التي تتطلب اهتماماً بشــأن الدور المســتقبلي وطبيعة دراســات البيانات الصغيرة، وتوســيع نطاق البيانات الصغيرة إلى أشكال مختلفة من البنى التحتيــة للبيانات والآثار المترتبة عليهــا، والعواقب الناتجة عن حفظ البيانات والوســاطة فيها على أســاس الربح. ومع ذلك، فمن الواضح أن التحليل النقدي للتغيرات التي تحدث ليست مواكبة للتطورات التقنية والتجارية، فالواقع أنه حتى الآن كان هناك عدد قليل من الانطباعات النقدية المبنية على أســاس التفكير من خلال الآليات والدلالات في كيفية إنتاج البيانات الصغيرة، وتوســيع نطاقها واســتخدامها، والتنظير لأدواتها - قواعد البيانات والبنى التحتية للبيانات - التي يتم اســتخدامها لتنظيم وتبادل وتحليــل البيانات. ونتيجة لذلك، هناك نــدرة في المعرفة المتعلقة بالأنظمــة الاجتماعية التقنية الناشــئة والمعقدة التي يتم إنشاؤها، مع عدد قليل من دراسات الحالة المعمقة. ومن ثم، يتضح أن هناك حاجة للكثير من البحث والتأمل لفهم البنى التحتية للبيانات في جميع أشكالها.





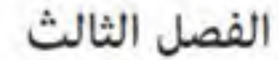
الفصل الثالث

البيانات المفتوحة (Open Data) والبيانات المرتبطة (Linked Data)

بالنظــر إلى التكاليف والنفقات والموارد المطلوبة لإنتاج مجموعات البيانات وقيمتها في الكشف عن معلومات عن العالم من حولنا، فقد تمّ بشكل عام تقييد الوصول إليها بطريقة أو بأخرى، فعلى ســبيل المثال يُتاح الوصول للمستخدمين المسجلين المعتمدين فقط، أو من خلال المطالبة برسوم اشتراك، أو عن طريق حصر الكيفية التي يمكن فيها استخدام البيانات بواسطة السياســات والتراخيص. وحتى عندما كانت مجموعات البيانات متاحة ومفتوحة نســبياً، فقد كانت تتطلب معدات وأدوات متخصصة لفهمهــا والوصول إليها، مثل أجهزة الحاســب، والبرمجيات، والمهارات المتعلقة بطرق الإحصاء ورسم الخرائط والمعرفة السياقية للحقل أو الموضوع الذي يجري التطرق إليه، مما يتعدّى بكثير قدرات عموم الناس. ونتيجة لذلك، كانت البيانات والمعلومات المستمدة منها تقليدياً مغلقة إلى حدّ كبير بطبيعتها - أي مقفل عليها داخل مؤسســة أو محفوظ أرشيفي. وفي الواقع، فقد مثل عدم إتاحة الوصول إلى مجموعات البيانات التي يمكن أن تساعد في الإجابة عن أسئلة معينة مصدر إحباط على مدى قرون للباحثين، والمحللين، والصحفيين، ومنظمات المجتمع المدني.

• ومن هذا المنطلق ســعت حركة البيانات المفتوحــة لتحويل هذا الوضع جذريًّا، ليس فقط من خلال فتح البيانات وإتاحتها لإعادة استخدامها على نطاق أوسع فقط، ولكن أيضاً من خلال توفير أدوات بحثية سهلة الاستخدام والتي لا تتطلب الحاجة إلى مهارات تحليلية متخصصــة. وبُنيت هــذه الحركة على ثلاثة مبــادئ هي: الانفتاح، والمشــاركة، والتعاون (البيت الأبيض White House 2009)؛ فمن خلال الشفافية، والمشاركة، والعمل معاً يمكن إدراك قيمــة البيانــات للمجتمع. وتهدف الحركة إلى تحرير القــدرة على إنتاج المعلومات والمعرفة، بدلاً من حصر إمكانية الاســتفادة من قوة البيانات على منتجيها أو أولئك الذين يمكنهم دفع ثمن إتاحة الوصول إلى البيانات. وقد تركز الاهتمام، على وجه الخصوص، على فتح البيانات التي تمّ إنتاجها من قبل أجهزة الدولة وغالباً ما يطلق عليها معلومات القطاع





العــام (Public Sector Information – PSI) أو البحوث المهولة من القطاع العام (انظر الفصل الثاني)، مع الأخذ بالحسبان بأن هذه البحوث قد تم تمويلها من قبل الخزينة العامة لفائــدة عامــة الناس، مع تركيز أقل على فتــح البيانات التي تمّ إنشــاؤها من قبل القطاع الخاص، والتي قد يكون لديها قيمة خاصة أكبر عند منتجيها على وجه الخصوص.

وقــد تطورت حركة البيانات المفتوحة خلال العقديــن الماضيين، بالتزامن مع - ولكن مع فصــل واضح إلى حدّ كبير عن - حركة الحــق في المعلومات (Right to Information-RTI) المعنية باتخاذ قرارات الشــفافية وصنع السياســات من قبل الوكالات والمعروفة أيضاً باســم حرية المعلومات (Freedom of Information)، وحركتي المصدر المفتوح والعلوم المفتوحة المعنيتين بإعادة تشكيل البرمجيات والنشر إلى أشكال مفتوحة وشفافة فيما يتعلق بالترخيص، وحقوق الطبع والنشر، وحقوق الملكية الفكرية. فمنذ آواخر ســنة 2000م، اكتســبت حركة البيانــات المفتوحــة مكانــة بارزة وزخم ثابت بشــكل ملحــوظ، ابتداء مــن حملة صحيفة الغارديــان في المملكة المتحدة تحت مُســمى" تخليص بياناتنــا- Free Our Data" (www. theguardian.com/technology/free-ourdata)، ودعــوة منظمــة التعــاون الاقتصادي والتنمية (Organization for Economic Cooperation and Development - OECD) حكومــات الدول الأعضــاء لفتح بياناتها في عام 2008 م، وإطلاق الحكومة الأمريكية في العام 2009 للموقــع الإلكــتروني (www.data.gov)، وهو موقع على شــبكة الإنترنت صُمّم لتوفير الوصــول إلى مجموعات البيانات غير الحساســة والتاريخية التي تحتفــظ بها أجهزة الدولة على مســتوى الولايات ومســتوى الحكومة الفيدرالية بالولايات المتحــدة الأمريكية، وتطوير البيانات المرتبطــة والترويج لـ الويب الــدلالي (Semantic Web) كعنصر موحّد لتكنولوجيا الانترنت في المســتقبل، والذي يُمكــن البيانات المفتوحة والبيانــات المرتبطة من أن تتصل في الغالب بشــكل طــردي (Berners-Lee 2009). ومنذ عــام 2010 م، حذت حذوها عشرات الــدول والمنظمات الدولية على ســبيل المثال، دول الاتحــاد الأوروبي (-European Union EU)، وبرنامــج الأمم المتحدة الإنمــائي (-United Nations Development Programme UNDP) مما جعل الآلاف من مجموعات البيانات المقيدة ســابقاً متاحة ومفتوحة بطبيعتها للاســتخدام غير التجاري والتجاري (انظر DataRemixed 2013). وقد أســهم في مثل هذا التحوّل في الموقف مجموعات الضغط الدولية والوطنية المؤثرة مثل مؤسسة المعرفة المفتوحة



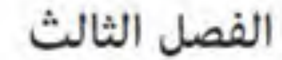
(Open Knowledge Foundation) ومؤسســة ضوء الشمس (Sunlight Foundation)، ورافقــه ضغــط من المجموعات والــشركات العاملة في مجال الاقتصاد المعــرفي، بالإضافة إلى كبار موظفي الخدمة المدنية المقتنعين بالحجج المستخدمة، وعشرات الجماعات المحلية التي تسعى للاستفادة من البيانات المدنية.

وفي حين يتم عرض حجج حركة البيانات المفتوحة بطريقة بديهية، باســتخدام مفردات الاســتعارة البلاغية مثل الشفافية، والمساءلة، والمشــاركة، والابتكار، والنمو الاقتصادي، فإن الانفتــاح السريع للبيانات الحكوميــة والعلمية لم يلق ترحيباً عالميــاً. فالواقع، أن العديد مــن الآثار الاجتماعية والسياســية والاقتصاديــة لفتح البيانات يجري حالياً الكشــف عنها ومناقشــتها وانتقادها. وعلاوة على ذلك، وللمفارقة إلى حــدّ ما، تعزّز نطاق ومدة حقوق الملكيــة الفكرية على مدى نصف القرن الماضي، بمعنى أن نمــو حركة البيانات المفتوحة تمّ بالتوازي مــع زيادة في حقوق الملكية الفكريــة (Pollock 2006)، فالمقاومة التي واجهتها فكرة توفير حرية الوصول إلى البيانات العلمية نوقشت في الفصل السابق، وسيتم في القسم الأخير من هذا الفصل مناقشة الانتقادات الموجهة لدعوى البيانات المفتوحة. فبداية، ومع ذلــك، يحدّد هذا الفصل خصائص البيانات المفتوحــة والبيانات المرتبطة، ومختلف الطرق التي هيأت لقضية البيانات المفتوحة، والاقتصاديات المرتبطة بتقديم بيانات مفتوحة، وعلى الرغــم مــن أن فتح البيانات الحكومية ما يزال جزئياً، فليس هنالك شــكّ في أن تحوّلاً كبيراً يحدث بشأن الكيفية التي تتم من خلالها عرض ومشاركة البيانات بشكل عام.

البيانات المفتوحة (Open Data):

عند أحد المســتويات، يُعد تعريف مصطلح البيانات المفتوحة بسيط نسبياً، فقد حدّده بولوك (2006) على سبيل المثال كالتالي: «تكون البيانات مفتوحة إذا كان لأي شخص الحرية في اســتخدامها، وإعادة استخدامها، وإعادة نشرها – فقط شرط موافقتها لمتطلب إسنادها و/ أو مشــاركتها على حدّ ســواء». ولكن هذا التعريف يُلمّح إلى الطبيعة المتنوعة لما يمكن أن يكون عليه المقصود من لفظة «الفتح» وكيف أنها يمكن أن تعني أشياء مختلفة بالنسبة لمختلــف الوكالات في ســياق حقوق الملكية الفكرية، فقد يشــير الانفتاح إلى الاســتخدام أو إعــادة الاســتخدام، أو إعادة الصياغــة، أو إعادة التوزيع، أو إعــادة النشر والتوزيع، أو





إعـادة البيع، وقد يكون لكلّ منها شروط وأحكام تختص بها، فعلى سـبيل المثال قد يكون للمستخدم القدرة على اسـتخدام مجموعة البيانات بحرية بموجب ترخيص، ولكن ذلك لا يعنـي إمكانية إعـادة التعامل مع هذه البيانات بهدف الربح التجـاري أو إعادة البيع، أو أي اسـتخدام قد يتطلب الإسـناد. وبعبارة أخرى، قد يكون الوصول إلى مجموعة البيانات مفتوحاً، ولكن ليس بالضرورة أن يكون كل ما يمكن فعله مع البيانات الممكن الوصول إليها مفتـوح أيضاً. وفي حـالات أخرى، قد تحتفظ المنظمات ببياناتهـا وتتحكم بإتاحة الوصول إليها بنفسها، ولكنها تتيح بحرية البيانات الوصفية المتعلقة بها فقط، أو قد تسمح المنظمة لبعض المستخدمين بالوصول إلى البيانات، ولكنها لا تمكنهم من إعادة نشرها إلا إذا كان ذلك سيحقق قيمة مضافة لها وبدون الوصول إلى البيانات الأولية الأساسية.

ومع ذلك، سـعت عدد من المنظمات إلى تحديد الخصائص المثالية للبيانات المفتوحة، على سبيل المثال يؤكد تعريف الانفتاح بأن العمل هو مفتوح إذا استوفت حالته في النشر الـشروط المنصوص عليها في الجدول رقم (٣-١)، والتي تفرض بعض القيود على الوصول، والاسـتخدام، وإعادة الصياغة، وإعادة التوزيع والنشر، وتشـجع كذلك بشكل نشط على الاستخدام الربحي للبيانات المفتوحة دون أي تعويض مالي لمنشئ البيانات الأصلي. وبالمثل، حدّدت منظمة الحوكمة المفتوحة (OpenGovData) تسـعة مبادئ للبيانات المفتوحة تتعـدى مجرد نمط التوزيع إلى النظر أيضاً في طبيعة البيانات نفسـها (انظر الجدول رقم 3-2). وبالإضافة إلى هاتين المجموعتين من الشروط / المبادئ، اقترحت وكالات أخرى مثل منظمة التعاون الاقتصادي والتنمية (2008) (OECD) والحكومة الأسترالية (Fitzgerald 2010) حاجـة البيانات المفتوحة إلى أن تكون مصحوبة بقوائم الأصول وآليات اكتشـاف البيانات، من أجل إصدار المؤشرات المتعلقة بجودة البيانات وموثوقيتها، واستخدام النماذج والمعايـير التي تتيح قابلية التشـغيل البيني للبيانات المفتوحة، وتوفير بنى تحتية سـهلة الاسـتخدام للبيانات لتسـهيل النشر المنتظم وإيجاد مجموعات بيانات جديدة، بالإضافة إلى عدد من الأدوات الأساسـية والتخصصية التي من شأنها أن تُحسن الاستخدام والتحليل للبيانات. ويُعد ضمان التشـغيل البيني مهماً بشـكل خاص لأنه يسـمح بربط مجموعات البيانات المختلفة، مما يُسـهم في بناء مجموعات بيانات جديدة أكثر تعقيداً واكتشاف رؤى جديدة (مؤسسة المعرفة المفتوحة Open Knowledge Foundation 2012). كما لا توجد قيـود فيما يتعلق بمجالات التركيـز للبيانات المفتوحة - حيث يمكـن أن تتعلق بأي مجال

من المجـالات الاجتماعية، والاقتصادية، والتجارية، والثقافية (وسـائل الإعلام والمكتبات والتراث)، والبيئية، أو الظواهر العلمية - ولكن، بشكل عام، ينصب التركيز حتى الآن على فتح البيانات التي تتصل بالسياسـات العامة عالية المستوى ولها قيمة تجارية عند إعادة استخدام مثل البيانات المتعلقة بالاقتصاد، والنقل، والبيانات الجغرافية.

جدول رقم (٣-١)

الخصائص المثالية للبيانات المفتوحة المتعلقة بتعريف الانفتاح

إتاحة الوصول Access	ينبغـي أن يكون العمل متاحاً بمجموعه ككل وبلا تكلفة إعادة إنتاج تتجاوز المعقول كما ينبغي أن تتوفر بشكل مناسب قابل للتعديل
إعادة التوزيع Redistribution	ينبغـي أن لا تُقيّـد الرخصـة أي طرف عن بيـع العمل أو توزيعه سـواء فيما يتعلق بالعمل نفسه أو كان ذلك العمل كجـزء من حزمة مركّبة من أعمال من عدّة مصادر مختلفة. وينبغـي أن لا يتطلب الترخيص إتاوة أو رسـوماً أخرى لهذا البيع أو التوزيع.
إعادة الاستخدام Reuse	ينبغي أن يسمح الترخيص بالتعديل على العمل وإجراء عمليات الاشتقاق منه ويسمح كذلك بتوزيع الأعمال المشتقة وفقاً لأحكام العمل الأصلي.
غياب القيود التقنية Absence of technological restrictions	ينبغـي تقديم العمل عـلى الصورة التي لا يوجـد معها عقبات تقنية تحول دون استخدامه أو إعادة توزيعه.
الإسناد Attribution	قـد يُلزم الترخيـص الإسـناد إلى المسـاهمين والمبدعين للعمل الأصلي كشرط لإعادة التوزيع وإعادة الاستخدام، وإذا تمّ فرض هذا الشرط فينبغي ألا تتطلب عملية الإسـناد إجراءات مرهقة ومُعقدة.



الصحة والنزاهة Integrity	من الممكن أن يشترط الترخيص حمل الإصدارات المعدّلة للبيانات أسماء تختلف عن العمل الأصلي كشرط لإعادة توزيعها.
عدم التمييز ضد الأشخاص أو المجموعات No discrimination against persons or groups	ينبغــي عدم تمييــز الترخيص ضــد أي شــخص أو مجموعة من الأشخاص وضمان إتاحة الوصول للجميع.
عدم التمييز ضد مجالات معينة في العمل No discrimination against fields of endeavor	ينبغي عدم تقييد أي شخص من الاستفادة من العمل في أي مجال معين. على سبيل المثال، لا ينبغي تقييد العمل من استخدامه في الأعمال التجارية.
توزيع الترخيص Distribution of licence	ينبغي تطبيــق الحقوق المتعلقة بالعمل عــلى جميع الذين يتم إعــادة توزيعه عليهم دون الحاجة لتنفيذ ترخيص إضافي من قبل تلك الأطراف.
عدم تعيين الترخيص على حزمة أعمال Licence must not be specific to a package	ينبغــي أن لا تعتمد الحقوق المتعلقة بالعمل على كونه جزءاً من حزمة معينة.
عدم تقييد الترخيص توزيع الأعمال الأخرى Licence must not restrict the distribution of other works	ينبغــي أن لا يضع الترخيص قيــوداً على الأعمال الأخرى التي يتم توزيعهــا جنباً إلى جنب مع العمل المرخص. على ســبيل المثال، ينبغــي أن لا يصّر الترخيص على أن تكون كل الأعمال المســتمدة منه مفتوحة أيضاً.

المصدر: مقتبس من http://opendefinition.org/od.

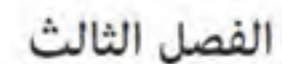


جدول رقم (٢-٣)

مبادئ منظمة الحوكمة المفتوحة للبيانات المفتوحة

يجب أن تكون البيانات كاملة Data must be complete	تكون كافة البيانات متاحة، مع مراعاة أحكام وشروط الخصوصية، والأمن أو قيود الصلاحيات.
يجب أن تكون البيانات أساسية (أولية) Data must be primary	تُنشر البيانات كما تمّ جمعها من المصدر، مع أفضل مستوى ممكن من التفاصيل، وليست بأشكال مجملة أو معدّلة.
يجب أن تكون البيانات في الوقت المناسب Data must be timely	إتاحة البيانات في أسرع وقت ممكن عند الحاجة لها للحفاظ على قيمة البيانات.
يجب أن تكون البيانات في المتناول Data must be accessible	إتاحة البيانات لأكبر عدد ممكن من المستخدمين ولأوسع مجموعة من الأغراض.
يجب أن تكون البيانات قابلة للمعالجة الآلية Data must be machine-processable	تنظم البيانات بشكل مناسب للسماح بالمعالجة الآلية لها
يجب عدم تمييز عملية الوصول Access must be non-discriminatory	تتوفر البيانات لأي أحد، مع عدم اشتراط التسجيل.
يجب أن تكون صيغ البيانات غير مخصصة الملكية Data formats must be non-proprietary	تتوفر البيانات في صيغة لا يكون فيها سيطرة حصرية عليها لأي كيان.





يجب أن يكون ترخيص البيانات مجانياً Data must be licence-free	لا تخضع البيانات لأية أحكام في حقوق التأليف والنشر وبراءات الاختراع والعلامات التجارية. وقد يُسمح ببعض قيود الخصوصية والأمن والصلاحيات المناسبة بحسب ما تفرضه التشريعات الأخرى.
يجب أن يكون الامتثال قابلاً للمراجعة وإعادة النظر Compliance must be reviewable	يجب أن يتم تعيين شخص كضابط اتصال للرد على الناس الذين يحاولون استخدام البيانات أو تقديم شكاوى حول انتهاكات المبادئ الرئيسية. ويجب أن يكون هناك هيئة أخرى ذات اختصاص لتحديد ما إذا تم تطبيق المبادئ بشكل مناسب.

المصدر: مقتبس من http://www.opengovdata.org

وإلى حدّ كبير، تعد هذه الخصائص المثالية البيانات المفتوحة مُنتجاً، وليست خدمة. وفي المقابـل، فقد جـادل Gurstein (2013) حول الحاجة إلى إعـادة التفكير بالبيانات المفتوحة كعملية خدمية (Service process)- بوصفها تفاعلاً وتنطوي على علاقة بين مزودي البيانات والمستخدم النهائي. فبالنسبة له، لا يكفي أن تكون البيانات ببساطة متاحة من قبل المنظمات لإعادة استخدامها، بل ينبغي أن يكون فتح البيانات موجّهاً أكثر باعتباره خدمة، مع الأخذ بعين الاعتبار حاجات وتوقعات المسـتخدمين النهائيين منها، فهذه الإعادة لصياغة المفاهيم تقترح على أولئك الذين ينتجون البيانات المفتوحة اتباع طرقاً مختلفة لإدارة البيانات والمسـتخدمين النهائيـين، كالآثار المترتبة عـلى قضايا إعادة تحديد المصدر الخاص بالبيانات. فبالنسـبة له، يتطلب هذا النوع من النهج مشـاركة المسـتخدمين النهائيين في التخطيط، والتطوير، وإدارة المشـاريع الجارية للبيانات المفتوحة، وكذلك استخدام مقاييس لتقييم مدى نجاح مثل هذه المشـاريع في تعزيز الصالح العام، وهذا يتطلب برامـج بناء القدرات الرامية إلى رفع مهارات المسـتخدمين ليكونوا قادرين على إدارة ومعالجة وتحليل البيانات بشـكل مناسـب وفعال. ورغـم أنه من السـهل أن نتفق عـلى أنه ينبغي تسـليم البيانات المفتوحـة كخدمة، إلا أن البيانات المفتوحة بطبيعتها لا تنتج دخلاً أو تنتج القليل مقابل تمويل هذه الترتيبات الخدمية فضـلاً عن التكاليف المتعلقة بفتح البيانات، فالواقع أن ذلك قد يكون من الطموح إلا إذا تم تطوير نماذج تمويل فعالة (كما نوقش أكثر بالكامل أدناه).

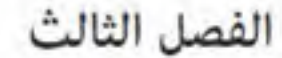


البيانات المرتبطة (Linked Data):

تتمثّل فكرة البيانات المرتبطة بتحويل شبكة الانترنت من شبكة الوثائق إلى شبكة البيانات من خلال إنشاء الشبكة الدلالية (Semantic web) (Berners-Lee 2009; P. Miller, 2010)، أو ما اصطلح عليه Goddard and Byrne (2010) بمصطلح «المقروءة آلياً على شبكة الإنترنت» (Machine-Readable Web). وتدرك هذه النظرة بأن كافة المعلومات التي يتم تبادلها على الإنترنت تحتوي على تنوع غني للبيانات - أسماء، وعناوين، وتفاصيل المنتج، والحقائق، والأرقام، إلى آخره. ومع ذلك، لا يتم بالضرورة تحديد هذه البيانات رسمياً على هذا النحو، كما أنها ليست منتظمة رسمياً في طريقة يمكن معها جمعها واستخدامها بسهولة. فالواقع، أن معظم وثائق شبكة الإنترنت غير منظّمة إلى حدّ كبير في طبيعتها. ومن خلال ترميز الوثائق وهيكلتها باستخدام معرّفات فريدة واستخدام لغة التوصيف (Markup Language)، يمكن جعل البيانات التي تحتويها هذه الوثائق مرئية، وتمكين الآخرين تلقائياً من دمجها، ومعالجتها، وفهمها، وربطها مع البيانات الأخرى ذات الصلة (P. Miller, 2010).

ومن أجل تمكين شبكة الإنترنت الدلالية من أن تعمل كبيانات مرتبطة، بيّن Berners-Lee (2009) حاجة الوثائق المنشورة على شبكة الإنترنت للمواءمة مع أربعة توقعات سلوكية. أولاً، يجب تعريف الكائنات والموارد ضمن كل وثيقة باستخدام لغة التوصيف الموسعة (XML- Extensible Markup Language) وتسميتها بشكل واضح لا لبس فيه باستخدام معرّفات الموارد الموحدة (Uniform Resource Identifiers -URIs). ثانياً، ينبغي استخدام الهيكل الضمني لشبكة الإنترنت لربط الوثائق والمعلومات (على سبيل المثال، استخدام معيار نقل النصوص التشعبية – أتش تي تي بي (HTTP) ومعرفات الموارد الموحدة (URIs) بحيث يمكن التنقيب في الأسماء والبحث فيها). ثالثاً، تسهيل اكتشاف المعلومات حول كائن مسمّى أو مورد (على سبيل المثال، عندما يقوم شخص بالبحث عن معرّف لكائن ما، تُقدّم له معلومات مفيدة باستخدام معايير إطار عمل وصف الموارد (Resource Description Framework -RDF). فإطار عمل وصف الموارد هو وثيقة تفصّل وتصف طبيعة معرّفات الموارد ضمن نطاق ما وينبغي أن يُحال كل معرّف إلى هذه الوثيقة لتحصيل معلوماته (Goddard and Byrne 2010). ويضمن استخدام هذه المعايير إطاراً مشتركاً لجمع البيانات عبر الإنترنت، بدلاً من اعتماد مجموعة كبيرة من





الأســاليب المتعارضة (Dietrich 2012). رابعاً، توفير وصلات للربط بين الكائنات والموارد ذات الصلة، إذا كانت معروفة، بحيث يمكن للمستخدمين اكتشاف المزيد من الأشياء.

وتمكّن هذه التوقعات الســلوكية من جعل الوثائق مقروءة آلياً ويمكن معالجتها بحيث يكون التعرف على طبيعة الكائنات فيها ممكناً (على سبيل المثال، الولايات المتحدة هي بلد، وباراك أوباما هو شخص)، وإزالة الغموض عن المفاهيم (على سبيل المثال، مارس هو قطعة شوكولاتة، والمريخ (مارس) هو كوكب؛ أو أن باراك أوباما هو مؤلف وثيقة مقابل أن باراك أوباما هو موضوع وثيقة)، وإنشاء روابط وعلاقات داخل البيانات (على سبيل المثال، باراك أوبامــا هو رئيس الولايــات المتحــدة) (Goddard and Byrne 2010). وبهذه الطريقة، يصبح من الممكن لوثائق شــبكة الإنترنت أن تكون قابلة للقراءة آلياً بحيث لا يقتصر عمل برنامج القراءة على فهم المحتوى ولكن يمكنه أن يســتمد أيضاً بيانات مشــتقة جديدة من خلال الاستدلال من المحتوى (Goddard and Byrne 2010).

وعندما تُنشر الوثائق بهذه الطريقة، يمكن اســتخلاص المعلومات على شــبكة الإنترنت وإعــادة تقديمها كبيانات وربطها بطرق لا حصر لها اعتماداً على الغرض. ولكن، وبحســب ما دوّن P. Miller (2010) فإن «البيانات المرتبطة يمكن أن تكون بيانات مفتوحة، ويمكن ربــط البيانات المفتوحة، ولكن مــن الممكن أيضاً للبيانات المرتبطــة أن تحمل تراخيصاً أو غيرها من القيود التي تحول دون اعتبارها مفتوحة»، أو إتاحة البيانات المفتوحة بطرق قد لا تجعل من السهل ربط بعضها ببعض. وبشكل عام، فإن أي وثائق مرتبطة ليست متوافرة على شــبكة الإنترنت أو تقع ضمن نطاق نظام اشــتراك مدفوع هــي أيضاً بيانات مفتوحة بطبيعتها. فبحســب ما يــرى Berners-Lee (2009)، ينبغي أن تكــون البيانات المفتوحة والبيانات المرتبطة متكافئة بشــكل مثالي، وإضافة لذلك فقد حدد خمســة مستويات لمثل هــذه البيانات، كل منها أكثر فائدة وقيمة تدريجيا (انظر الجدول رقم 3-3). وكان تطلعه إلى ما أســماه بيانات الخمســة نجوم (المستوى الخامس) – شــبكة دلالية عاملة بالكامل. ولكــن تجدر الإشــارة إلى أنه بالنســبة إلى العديد من المنظمات، فــإن مجرد الحصول على مستويات أعلى من المستوى الأول يُعدّ إنجازاً، وكون المستوى الخامس بعيد المنال عنها دون عمالة ماهرة، وموارد إضافية، وأدوات جديدة.



الجدول (٣-٣)

المستويات الخمسة للبيانات المفتوحة والبيانات المرتبطة

المستوى	الصورة / الشكل	الفوائد	التكاليف
1	غير مقروءة آلياً	البيانات متوافرة.	يتم قفل البيانات في الوثيقة ويصعب تحريرها لاحقاً.
2	مقروءة آلياً ولكن بصيغة حصرية (على سبيل المثال، ملف إكسل).	يمكن تحليل البيانات باستخدام البرمجيات الحصرية / الاحتكارية؛ كما يمكن تصديرها إلى صيغ أخرى.	تعتمد على برامج حصرية للوصول إليها واستخدامها.
3	مقروءة آلياً باستخدام صيغ عامة غير حصرية (على سبيل المثال، صيغة سي إس في- CSV)	يمكن تحليل البيانات باستخدام أية حزمة برمجية.	هل البيانات على الشبكة، ليست البيانات على الشبكة، وليست مرتبطة بطبيعتها، ومن ثم هي معزولة.
4	مقروءة آلياً باستخدام صيغ عامة غير حصرية واستخدام المعرفات ووصف الموارد (URIs and RDF)	يمكن الوصول إلى البيانات من أي مكان على شبكة الإنترنت، ومن السهل ربطها جنباً إلى جنب مع غيرها من البيانات، وإمكانية وصلها بالأدوات والمكتبات القائمة.	يمكن أن تزيد زمن تحضير البيانات وإدارة البيانات وتبادلها.
5	مقروءة آلياً باستخدام صيغ عامة غير حصرية واستخدام المعرفات ووصف الموارد ومرتبطة بالبيانات والبيانات الوصفية الأخرى	كما في المستوى الرابع، ولكن تصبح البيانات أكثر اكتشافاً وتمنح المستخدمين وصولاً كاملاً إلى مخطط (أنطولوجيا) البيانات	تحتاج إدارة فعالة للبيانات للحفاظ على الروابط الداخلة والخارجة.

المصدر: معدّلة عن http://sstardata.inio بواسطة Michael Hausenblas.



الدَّعْوَى نحو البيانات المفتوحة:

في حـين كان لـدى بعض البلدان، مثـل الولايات المتحدة، تقليـد عريق في جعل بعض بيانـات القطاع العـام متاحة بحرية، وخاصـة البيانات ذات الفائدة العاليـة مثل بيانات التعداد، ورسم الخرائط، والطقس، فإن إتاحة الوصول، في بعض البلدان الأخرى، يتمّ بشكل عـام من خلال نمـوذج تغطية التكاليف الذي يتيح البيانات بمقابـل مالي وبموجب اتفاقية ترخيص معتمدة. ففي المملكة المتحدة، تخضع الوثائق الحكومية لحقوق التأليف والنشر الملكيـة (Crown Copyright) وتُحكم البيانات ذات الفائدة العالية بواسـطة الصناديق التجارية (على سـبيل المثال، بيانات الخرائط ضمن هيئة المسح الجغرافية، وبيانات أحوال الطقس في مكتب الأرصاد الجوية) التي تعمل بصفة كيانات احتكارية، وتتحكم في الوصول إلى مجموعات بيانات رئيسية (Arthur and Cross 2006; Pollock 2006). وعلى مدى العقديـن الماضيين، تمّ تجميـع مجموعة من الحُجج لبناء دَعْوَى إعـادة صياغة دور هذه الصناديـق وفتح بيانات القطاع العام للجمهور. وبصفة عامة، يمكن تقسـيم هذه الحُجج والاسـتحقاقات في خمسـة أشـكال، والتي ينظر إليها على أنها ذات فائدة لكل من حامل البيانات والمستخدم النهائي.

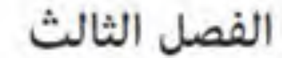
أولاً: حيث أن الكثير من البيانات في إطار الهيئات العامة والمنظمات غير الحكومية تتعلّق بالعمليات التشـغيلية لتلك الهيئات، فهي توفّر وسـيلة يمكن من خلالها قياس مدى نجاح مختلف البرامج والأنشـطة الخاصة بها. ولـذا فإن فتح هذه البيانات للتدقيق والرقابـة العامة سـيجعل أعمال المنظمة وصنع القرار فيها تتمتع بالشـفافية ويمكن اسـتخدامه لتعزيز المسـاءلة وتقييم القيمة الناتجة مقابل المال المستثمر (Janssen 2012; Gordon 2013). وقد تزايدت أهمية هذه الشـفافية والمساءلة في الخطاب العام في عصر سياسـات التقشـف ومحدودية الموارد، وفي سـياق المناقشات الجارية حول المشتريات، والهدر، والعجز الظاهر في الخدمات العامة.

ثانياً: يسـمح إتاحة الوصول للمسـتخدمين النهائيين لبيانات المنظمة، على زعم البعض، برفع مسـتوى الوعي حيال القضايا المثارة، مما يسـهّل الاختيـار وصنع القرار فيما يتعلق بالخدمات العامة، ويشـجع على المشـاركة الفاعلة والواعية في المجال العام

(Janssen 2012; Yiu 2012). وعلاوة على ذلك، فإن إتاحة الوصول للمستخدمين النهائيين لبيانات المنظمة يعزّز المواطنة الفاعلة والمشاركة السياسية في تشكيل كيفية أداء وتنظيم الحكم المحلّي (Huijboom and Van der Broek 2011). وهذه، بدورها، تشجّع الابتكار الاجتماعي، وتعمل على تعزيز العلاقات بين أفراد المجتمع، ورفع مستوى النقاش العام (Yiu 2012)، ونتيجة لذلك، تمكّن البيانات المفتوحة من بناء الديموقراطية القائمة على المشاركة.

ثالثاً: يشجّع فتح البيانات حول المنظمة وأدائها على استخدامها هذه البيانات للاستفادة في تحسين الكفاءة التشغيلية والإنتاجية من خلال المراقبة واتخاذ القرارات بالأدلة والبيّنات المثبتة. وعلاوة على ذلك، تحصل وحدات الأعمال داخل المؤسسة على حق الوصول إلى البيانات وإمكانية استخدامها عبر المنظمة بكاملها، مما يؤدي إلى رؤى ومعارف جديدة ومزيد من التفكير المشترك والكفاءة (Northcutt 2012; Verwayen et al. 2011). وإضافة لذلك، فإنه يمكنهم اكتساب ملاحظات وردود قيّمة وتلقّي النصح والمشورة من الوكالات الخارجية التي تقوم باستخدامها، وتحليلها، وتفسيرها، مما يرفع من مستوى الجودة، والنزاهة، والفائدة من البيانات. وتؤدي هذه التحليلات الداخلية والخارجية إلى تعزيز الحوكمة التنظيمية وإدارة المجتمع على نطاق أوسع، على سبيل المثال التصدي للغش والتزوير والجرائم الأخرى من خلال إنشاء خدمات أكثر فعالية (Huijboom and Van der Broek 2011). كما أنها تمكن من الرصد والتقييم المستمرين لأثر السياسات والبرامج الجديدة (مؤسسة المعرفة المفتوحة Open Knowledge Foundation 2012).

رابعاً: يُمكّن توفير البيانات المفتوحة من إثراء العلامة المميزة، فإتاحة البيانات بحرية يرفع من شأن المنظمة، ويدلّل على ريادتها، وروح المبادرة فيها، ويخدم مهمتها العامة، ويزيد مستوى الاتصال والتفاعل مع العملاء، والمستخدمين النهائيين ويقود التوجه إلى الخدمات المؤسسية (Verwayen et al. 2011)، وفي المقابل، يبني هذا الثقة والسمعة، ويعمل على تكوين انطباع عن المعايير الفاضلة للمؤسسة.

خامساً: بينما تمتلك بعض مجموعات البيانات المنشئة والقائمة في العلن قيمة اقتصادية هامشية، على سبيل المثال تلك المتعلقة بالتراث الثقافي، فإن الكثير من مجموعات البيانات الأخرى لديها قيمة تجارية كبيرة. ويمكن استخدام البيانات التي تمّ إنشاؤها علناً لإضافة قيمة إلى البيانات التجارية القائمة، وإنشاء تطبيقات وخدمات جديدة ومن ثم أسواقاً جديدة، وتحسين المعرفة واتخاذ القرارات للأعمال (Janssen 2012; Yiu 2012). ولذا فليس من المستغرب عندئذ حرص المصالح الصناعية على الدعوة لفتح البيانات المُنتجة علناً لإعادة استخدامها في القطاع التجاري، وخاصة البيانات التي طالما عُرفت باحتوائها وإمكانية إنشائها لقيمة كتلك التي تُدار من قبل صناديق تداول القطاع العام (على سبيل المثال، بيانات الخرائط). فهناك من يزعم بأن تقييد الوصول إلى البيانات العامة، والتي تمّ إنشاؤها من أجل المصلحة العامة، يخنق الابتكار ويعمل على فقدان وزن مميت للاقتصاد؛ أي أن حصر استخدامها فقط لمن لهم القدرة على تحمّل تكلفتها، هو تسعير لإخراج الآخرين من السوق (Pollock 2006; Yiu 2012). ولذا فإن الاستثمار العام في إنتاج البيانات لم يستغل بالشكل المناسب كما تم فقدان العديد من الفرص التي من شأنها أن تضفي قيمة على نشاطات القطاع الخاص (Yiu 2012). وإضافة إلى ذلك، فُقد تأثير الارتفاع على الاقتصاد الشامل نتيجة فقدان القيمة الناتجة عن فتح البيانات العامة (Northcutt 2012).

وتتفاوت كيفية احتساب أوزان هذه الحجج الخمس عندما يتم تقديمها عبر السلطات القضائية والسياقات المختلفة؛ ومع ذلك، فهي تشكّل مجتمعة نظاماً ذا منطق قوي حقّق نجاحاً ملحوظاً في إقناع المنظمات لفتح البيانات الخاصة بها، لأسباب ليس أقلها أنها تتوافق مع العمليات الجديدة للخصخصة (Newliberalism) التي تعزز تسويق الخدمات العامة (Bates 2012). وفي الواقع، فإن المحرك للحجج المتعلقة بفتح البيانات في الغالب هي المصالح السياسية والتجارية التي تهدف إلى دفع المؤسسات العامة والمنظمات غير الحكومية للحصول على وصول إلى سلعة ثمينة، بدلاً من السعي لتحدي مفهوم الملكية الفكرية في حدّ ذاتها (Verwayen et al. 2011).

في المقابل، ومما لا يثير الدهشة في هذا السياق، أن الحجج والدعاوى لقطاع الأعمال لإتاحة بياناتها بحرية لإعادة استخدامها على نطاق أوسع لم يجر تطويرها والترويج لها

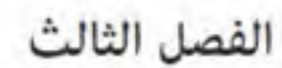


على نحو كاف. ومع ذلك، فإن العديد من الحجج نفسـها لفتح بيانات القطاع العام تصلح لبيانـات الأعمال أيضاً، أو على الأقل عناصر مختـارة من هذه البيانات. ففتح البيانات من شـأنه أن يمكن الشركات من الاسـتفادة من التحليل والرؤى التي تنتج بواسـطة الغير من المسـتخدمين النهائيين، والجمع بين البيانـات ومجاميع البيانات الأخـرى التي تعمل على إثرائهـم بطرق مثمرة، وتعزيـز التعاون مع الشركاء والموردين مـما يصنع فعالية في مجال الخدمات اللوجستية والمنتجات الجديدة، وبناء علاقة مع العملاء تساعد على توسيع قاعدة العمـلاء وتعزيز الملف التعريفي للشركة وسـمعتها (Deloitte 2012). وعلاوة على ذلك، فإنها قد تسـاعد على تطويـر الشراكات بين القطاعين العام والخاص، وفي هذا السـياق تمّ اعتـماد نهج الابتكار المفتوح، والذي حقق نجاحاً فيما يتعلق بالبرمجيات المفتوحة، والعمل على مبدأ أن تقاسـم الموارد سـينتج المزيد من القيمة التجارية دون الحرص على اكتنازها بغـيرة والتحـرز عليها. ففي الحالات التي تلقت الشركات الخاصة أموالاً عامة للمسـاعدة في البحـث والتطوير، يمكن إقامة دَعْوَى عـلى أن أي بيانات تنتج عن مثل هذه المحاولات ينبغي أن تكون مفتوحة إما بشـكل كامل وإما بشـكل انتقائي كوسـيلة لإنتاج سلعة عامة لمثل هذا الاستثمار.

اقتصاديات البيانات المفتوحة:

هنـاك عدد من العوامل الكابحة / المثبطة لفتـح البيانات، فبعض هذه العوامل تتصل بالمخاوف المتعلقة بقضايا الضمان والمسؤولية، والمخاوف بشأن الخصوصية والأمن، والحرج المحتمـل على جودة البيانـات أو مداها، والكفـاءة الفنية لبناء واجهات مناسـبة لبرمجة التطبيقـات (Applications Programming Interfaces- APIs) وإدارة البنية التحتية. وحتـى الآن، فالأمر الأكثر إثارة للقلق، مع ذلك، هو الشـأن المالي. فالبيانات قد تكون غير متنازعـة بطبيعتها، مما يعني أنه من الممكن توزيعها بتكلفة هامشـية، نظرياً على الأقل، ولكن ينبغي الدفع مقابل النسـخة الأولية جنباً إلى جنب مع إدارة البيانات القائمة وخدمة العمـلاء (Pollock 2006). وعـلى هذا النحو، قـد تمثل البيانات المفتوحة مـورداً مجانياً للمستخدمين النهائيين، ولكن إنتاجها ومشاركتها والمحافظة عليها هي بالتأكيد ليست دون تكلفـة معتبرة (خصوصاً فيما يتعلق بالتقنيات المناسـبة والموظفين المهرة). ففي كثير من





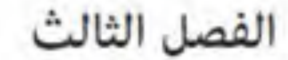
الحالات، كانت هذه البيانات أيضاً مصدراً رئيسـياً للدخـل للمنظمات، وفي حالة الشركات ميزة تنافسـية. زمن ثم يرتكز السـؤال الأساسي حول كيفية تمويل مشـاريع فتح البيانات على نحو مسـتدام في ظل عدم وجود مصادر الدخـل المباشر لها. ولذا هناك وجهان لهذه المعضلة في التمويل: فمن ناحية هناك التمويل اللازم من قبل الدولة لإنتاج البيانات وجعلها مفتوحة. ومن ناحية أخرى هناك التمويل للحفاظ على المبادرات التي يقودها المواطنون، والذي يعتمد على العمل التطوعي والمنح، ونماذج الأعمال التي من شـأنها تمكين الشركات التي تستخدم البيانات المفتوحة من الازدهار والتطور.

ويختلف تمويل خدمات البيانات الحكومية بين البلدان والوكالات، ففي كثير من الحالات، تُفـرض الضرائب لإنتاج ومعالجة البيانات. ولكن في بعض النظم القانونية والحالات، تكون خدمات البيانات معقدة بسـبب أربعة عوامـل: الأول، أن إجراءها يتم بالتعاقد عليها مع أطـراف ثالثة لإدارتها وتشـغيلها بالنيابة عـن الدولة، حيث يضيف الطـرف الثالث قيمة مخصصة احتكارية أو يجعل البيانات متاحة مقابل رسـم مالي. وقد حدث هذا مؤخراً مع الرمـوز البريدية الأيرلندية المرتقبة التي سـيتم إدارتها من قبل إحـدى الشركات نيابة عن الدولة وستمول هذه العملية من خلال بيع وترخيص البيانات. الثاني، أن البائعين الآخرين يتكتلون بنشاط في سبيل وقف تقديم البيانات المفتوحة لأنه يدمر نموذج أعمالهم. ثالثاً، أن بعض أجهزة الدولة تعمل باعتبارها صناديق تداول تجارية، فهي لا تحصل على كل تمويلها من عائدات الضرائب، ولكن جزءاً كبيراً من دخلها قائم على بيع البيانات. فهيئة المسـاحة الأيرلندية، على سبيل المثال، تعمل بهذه الطريقة كون أقل من نصف دخلها يأتي مباشرة من الدولة في شكل إعانة، ومن المسلّم به أن بعض هذه الدفعات التي تتلقاها تأتي من وكالات حكومية أخرى، ولكنها تأتي أيضاً من المؤسسـات الخاصة وعمليات الشراء الفردية. فإتاحة جميـع البيانـات الخاصة بها مجاناً يقـوّض قدرتها على العمل وتمويـل الخدمات الجارية لهـا. رابعـاً، أن جعل البيانات مفتوحة لا يتمثل بنشرها في شـكل محـدّد من قبل الدولة فقـط. فالكثـير من البيانات تحتاج إلى إعـادة توظيفها وتنسـيقها لتمكينها من أن تكون مفتوحة (على سـبيل المثال، إبهام المصدر لهـا، وتجميعها) ووضع أنظمة جديدة في المكان المناسـب لتحقيق هذا الهدف. وهذه ليسـت عملية هيّنة، وفي أوقات التقشـف وخفض النفقـات يعني هذا إعـادة تخصيص التمويل لتغطية ثمن هـذا العمل، الذي هو ضروري

أيضــاً للخدمات الأساســية. وتعني هــذه التعقيدات الأربعــة أن التصريحات، في كثير من الأحيــان، بــأن جميع بيانات الدولة قد تمّ بالفعل تغطيــة كُلَفها من خلال فرض الضرائب، وأنه ينبغي أن تكون متاحة بحرية، هي ليســت بســيطة كما يُراد لها أن تكون. أما فيما يتعلــق بالمبادرات التي يقودها المواطنون، فهذه أيضاً لديها تكاليف فيما يتعلق بالتوظيف والمعــدات والخدمات إذا أُريد لها أن تكون مســتدامة على المــدى الطويل. فكيفية تأمين هذه الموارد خارج نطاق العمل التطوعي والهدايا تُشــكل تحدياً صعباً. حيث تعتمد هذه المنظــمات، في الوقت الحاضر، إلى حد كبير على التبرعــات الخيرية وتمويل الدولة، والتي لا يحمــل أيّاً منها ضمانات للتجديد والاســتمرار. ومن ثم، فمن النتائج المحتملة المترتبة على التخفيضــات أو التقلبات في القاعدة المالية لخدمات البيانات المفتوحة الانخفاض في جودة البيانات، وبط ء الاستجابة، والابتكار، والأداء العام (Pollock 2006).

وعموماً يُجادل الدعاة للبيانات المفتوحة بأن الخيار الأفضل لتأمين قاعدة مالية مستقرة للبيانــات المفتوحة من داخل الدولة وخارجها هوعن طريــق الإعانات الحكومية المباشرة لتكاليفها. ويرى أنصار هذا النهج أن زيادة الإنفاق العام يعوّض في أربع طرق: الأولى، أن إتاحــة الوصول المباشر إلى البيانــات يحدّ من بعض التكاليف التي يتكبّدها المنتجون، مثل التوظيف المطلوب للتســويق، والمبيعات، والتواصل مع العمــلاء، ومراقبة شروط وأحكام الترخيــص (Pollock 2006). الثانيــة، أنــه يمكن للنموذج المفتوح الاســتفادة من العمل الإضافي الحر والابتكار من جموع المســتخدمين التي تضيــف قيمة معتبرة إلى مجموعات البيانات وللمنظمات من ناحية جودة البيانات، وتحليلها، والمعرفة المســتمدة، والمنتجات الجديــدة والابتكارات، والعلاقات والشراكات التجارية الجديدة (de Vries et al. 2011;) Houghton 2011). الثالثة، أن البيانات المفتوحة تنتج قيم فائضة متنوعة للمستهلكين، كما تولّد سلعاً عامة مهمة تستحق استثمار النفقات العامة فيها (Pollock 2009). الرابعة، أن البيانات المفتوحة تؤدي إلى منتجات مبتكرة جديدة من شأنها إنشاء أسواق جديدة، وهذا بدوره ســوف ينتج مداخيل للشركات وإيرادات ضريبية إضافية. وستكون هذه الإيرادات الضريبية فائضة عن التكاليف الحكومية الإضافية في فتح البيانات، وبعبارة أخرى فإن نهج التكلفة الصفرية أو الهامشــية هي أكثر فائدة على المدى الطويل من إستراتيجيات تغطية التكاليف واستردادها (المفوضية الأوروبية European Commission 2012).





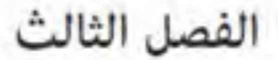
ولذا، يُعتقد أن فتح البيانات ينطوي على إمكانية إيجاد عدة عشرات المليارات من اليورو سنوياً في منطقة الاتحاد الأوروبي وحدها (مؤسسة المعرفة المفتوحة Open Knowledge Foundation 2012). ومن الأمثلة المستخدمة لدعم هذا الاعتقاد هو قرار الولايات المتحدة بإتاحة بيانات نظام تحديد المواقع العالمية والطقس للجمهور بحرية (Pollock 2006; de Vries et al. 2011). حيث يدعم كل منها صناعات بمليارات الدولارات والتي تولّد إيرادات ضريبية كبيرة وتوظّف أعداداً كبيرة من العمال. وقد قدّرت الإيكونوميست (The Economist) في العام 2013م بأن ثلاثة ملايين وظيفة في الولايات المتحدة وحدها تعتمد على نظام تحديد المواقع. فلو أن أحداً قارن استخدام بيانات الطقس العامة في الولايات المتحدة وأوروبا، فهناك فرق شاسع في استخدامها وإنتاج قيمة مضافة عنها، حيث يشير Pollock (2006) إلى أنه في حين " تساوي الاقتصادين تقريباً في الحجم، فإن صناعة أحوال الطقس التجارية في الولايات المتحدة هي أكبر بعشر مرات من نظيرتها في أوروبا كما أن صناعة إدارة مخاطر الطقس الوليدة هي أكبر بمائة مرة منها في أوروبا". وبعبارة أخرى، تمّ استخدام البيانات العامة المتاحة علناً بالمجّان أو بقيمتها الهامشية بطرق أنتجت عنها قيمة إضافية كبيرة. فأسواق البيانات المفتوحة تتعلق على حد سواء بأغراض مرتفعة (وهي قاعدة صغيرة من العملاء الذين يدفعون رسوماً كبيرة لقيمة مضافة عالية؛ على سبيل المثال، المستهدفين بشكل كبير بالتنبؤات الجوية)، وأغراض منخفضة (وهي تستهدف أعداداً كبيرة من المستخدمين طلباً لقيمة مضافة منخفضة حيث الخدمة مجانية، مدعومة من إيرادات الإعلانات؛ على سبيل المثال، تطبيقات حركة المرور) (de Vries et al. 2011). ويتضح التأثير المحتمل لفتح بيانات القطاع العام في أوروبا من خلال دراسة أجريت على واحد وعشرين مشروعاً للبيانات المفتوحة في عشرة بلدان والتي وجدت أن الانتقال إلى التكلفة الصفرية أو الهامشية لتكلفة العمليات التشغيلية زادت عدد المستخدمين ما بين ألف إلى عشرة آلاف في المائة، كما جذبت العديد من الأنواع الجديدة من المستخدمين، ولا سيما الشركات الصغيرة والمتوسطة (de Vries et al. 2011). وهذا يفصح عن أنه على الرغم من أن هناك بعض المكاسب الكبيرة، فقد بيّن Eaves (2013) أن غالبية الإيرادات الناشئة ستنتج عن سلسلة طويلة من الادخار والكفاءة والزيادة المتواضعة من التدوير.



وعلى الرغم من هذه الحجج، لم يصدر عن هيئة المحلفين حكم قاطع حول ما إذا كان فتح جميع بيانات القطاع العام ممكناً اقتصادياً وبشكل مستدام، وخاصة في المدى القصير. فهنالك بالتأكيد حالة تتعلق بأن البيانات المفتوحة من الممكن تسويقها، مما يؤدي إلى منتجات عالية ومنخفضة المستوى والتي يمكن بالتأكيد أن تنتج مصادر للدخل وفرص العمل. ومع ذلك، وكما بيّن de Vries et al. (2011)، فإن متوسط ما يكسبه مطورو التطبيقات من مبيعاتها سنوياً ليس سوى ثلاثة آلاف دولار فقط، مع الأخذ بالحسبان أن ثمانين بالمائة من تطبيقات الأندرويد المدفوعة يتمّ تنزيلها أقل من مئة مرة. وبالإضافة إلى ذلك، فقد لاحظوا أنه حتى التطبيقات الناجحة مثل ماي سيتي واي (MyCityWay) التي تمّ تحميلها أربعين مليون مرة، لم تنتج أرباحاً بعد. وبدلاً من ذلك، يستثمر أصحاب رؤوس الأموال في مشاريع محتملة الربح مع الحاجة إلى نموذج مستدام للأعمال. وقد يكون من الجيد أن يستغرق الأمر وقتاً لتطوير ابتكارات وأسواق جديدة؛ فعلى سبيل المثال، استغرقت الصناعات القائمة على نظم تحديد المواقع سنوات عديدة لتزدهر وتنضج بعد قرار جعل البيانات متاحة علناً والذي اتخذ في العام 1984م. كما قد يكون الحال أن بعض البيانات لها قيمة كبيرة جداً ولكن دون جدوى اقتصادية، مما يستوجب عادة عدم استمرار إعانتها مالياً إذا أريد لها أن تبقى مفتوحة بطبيعتها.

وفي ظل غياب طرح الاكتتابات الحكومية العامة لجميع التكاليف المتعلقة بفتح البيانات الخاصة بها، يجري النظر في حلول تمويلية محتملة أخرى ولا سيما نماذج الأعمال المختلفة. فقد بين Ferro and Osella (2013) ثمانية نماذج مختلفة لتستخدم، أو قد تستخدم لتمويل مبادرات البيانات المفتوحة (انظر الجدول ٤-٣). ولا تستبعد بعض هذه النماذج المختلفة بعضاً. في حين تعمل نماذج شبه - مفتوحة أخرى على توفير البيانات مجاناً لإعادة الاستخدام غير التجاري، وإلزام معيدي الاستخدام للأغراض الربحية بدفع مقابل مادي، أو الدخول في شراكات بين القطاعين العام والخاص حيث يزوّد القطاع العام البيانات وتقدّم الشركات الخاصة التمويل والخدمات المضافة لإتاحة الوصول وحقوق إعادة الاستخدام (OECD 2008). وفي نموذج آخر، يمكن بناء تحالف يمتلك مجموعة البيانات، وفرق العمل، والموارد، والأدوات، ويسهل بناء القدرات، ولكن يتقاضى رسوم العضوية من أعضاء التحالف لتغطية الخدمات ذات القيمة المضافة المشتركة. ويتوقف اعتماد أي نموذج، أو





مجموعة من النماذج، على المجموعة المســتهدفة لتحمّل تكلفــة إنتاج وصيانة البيانات – المستخدمين، المحدّثين / البائعين او الحكومة - وإذا كانت المنظمة تسعى لاسترداد التكاليف كاملة أو جزء هامشياً منها، أو حتى قيمة أكبر من التكاليف الكاملة التي يمكن استثمارها مرة أخرى في الخدمة (Pollock 2009; Ferro and Osella 2013).

الجدول (٤-٣)

نماذج التمويل لمبادرات البيانات المفتوحة

النموذج	الوصف
المُنتج المميز / خدمة Premium product / service	تقدّم للمستخدمين النهائيين منتجاً عالياً المستوى أو الخدمات التي تضيف قيمة إلى البيانات (على سبيل المثال، البيانات المشتقة، الأدوات والتحليل) بمقابل مادي، كدفعة مقطوعة ثابتة، أو رسوماً متكررة، أو الدفع - مقابل - الاستخدام، ودون استخدام حقوق احتكارية. وهذا يتيح لمنتج البيانات الحصول على مزايا المبادرة والسبق في مجال تسويق وبيع السلع التكميلية.
المنتج المجاني / خدمة Freemium product / service	تقدّم للمستخدمين النهائيين مجموعة متدرجة من الخيارات، متضمناً ذلك خياراً مجانياً يشمل عناصر أساسية (على سبيل المثال، ميزات محدودة أو عينات من مجموعات البيانات)، مع خيارات أكثر تقدماً تحقق قيمة مضافة بمقابل رسم مالي. تفتح المنتج / الخدمة على سوق أكبر اتساعاً وأكثر شعبية والمزيد من الاستخدام العرضي الممتد، مع الاحتفاظ بالخدمات / المنتجات المدفوعة، عالية المستوى، المتاحة للمستخدمين الأكثر تخصصاً.
المنتجات مفتوحة المصدر Open source	تقدّم منتجات / خدمات للمستخدمين النهائيين مجاناً دون مقابل، حيث تُدعم من خلال سبل التمويل الأساسية الأخرى أو غيرها من المنتجات / الخدمات.



نموذج الهدية الترويجية (البنى التحتية على نسق ماكنة الحلاقة والشفرات) Infrastructural razor and blades	يعرض المنتج او الخدمة ابتداء بالمجان أو بقيمة مخفَّضة (ماكنة الحلاقة) مما يشجّع على المتابعة بالاستخدام المدفوع مقابل (الشفرات). وقد يكون إتاحة الوصول إلى واجهات برمجة التطبيقات بالمجان، ولكن يكلّف هذا الاستخدام المحوسب المستخدمين على نموذج دفع المستحقات أولاً بأول (بحسب الاستخدام)، مع دعم هذا النموذج الأخير (الشفرات) للنموذج الأول (ماكنة الحلاقة).
المنصة الموجهة بالطلب Demand-oriented platform	تقدّم خدمات ذات قيمة مضافة عن طريق تنظيف، وتكرير، وتوحيد، وربط البيانات مما يوفّر مجموعة قياسية من الأدوات التحليلية، وإتاحتها من خلال متجر شامل (متجر الشباك الواحد) وتحميل المستخدمين التكلفة من خلال التسعير المجاني او المميز (تتصاعد تدريجياً بخيارات 1 و2 لاكتساب الكفاءة).
المنصة الموجهة بالتزويد Supply-oriented platform	يُعطى مطورو البرمجيات حرية الوصول إلى البيانات لإنشاء الخدمات والمنتجات التي يمكنهم بيعها إلى هيئات القطاع العام، والتي يمكنها بدورها أن تجعلها متاحة للآخرين من خلال نموذج المنتجات المجانية.
التوزيع المجاني مع الإعلان \Free with advertising	تُقدّم المنتجات / الخدمات مجاناً، ولكن يتلقى المستخدمون إعلاناً عند استخدام المنتج / الخدمة (المدرّة للدخل) أو تقدّم المنتجات / الخدمات من قبل شركات مختلفة وتوصف بالعلامات التجارية لها لتشجيع استخدام منتجاتها وخدماتها الأخرى (الدعم المتبادل).
تطوير العلامة البيضاء White-label development	يتم إنشاء منتج / خدمة مخصصة للعميل وتخصّص لاستخدامه الخاص، مع دفع العميل رسوماً لمرة واحدة أو اشتراك يتضمن تكاليف الصيانة والتحديث.

المصدر: جمعت من Ferro and Osella (2013) .



المخاوف المتعلقة بالبيانات المفتوحة:

باستثناء الكيفية اللازمة لتمويل مستدام لمبادرات البيانات المفتوحة، والخسائر المحتملة للشركات التي قامت ببناء نماذج أعمال قائمة على إعادة بيع بيانات القطاع العام (de Vries et al. 2011)، سيظهر للوهلة الأولى أن هناك عدداً قليلاً من السلبيات الناتجة عن فتح البيانات العامة للجمهور لإعادة استخدامها، وأن هناك العديد من الإيجابيات. ومع ذلك، فقد بدأ في الظهور مؤخراً عدد من الانتقادات المضادة التي تزعم أن حركة البيانات المفتوحة ليست محمودة من الناحية السياسية أو الاقتصادية، حيث إن بعض العناصر زائفة في أهدافها، وعلاوة على ذلك، أن هناك عدداً من التبعات الضارة التي يمكن أن تنجم عن فتح البيانات ويتضرر منها بعض المواطنين. ولا توحي هذه الانتقادات بفكرة التخلي عن التوجه نحو فتح البيانات، ولكن تؤكّد على وجوب أن تكون مبادرات البيانات المفتوحة أكثر وعياً وتعقّلاً في إتاحة البيانات، وكيفية استخدامها، وطرق تمويلها. ويمكن تقسيم هذه الانتقادات إلى ثلاث فئات رئيسية: تسهيل البيانات المفتوحة للتوجه للخصخصة (تأييد الرأسمالية المطلقة وعدم تدخل الدولة في الاقتصاد) والهيكلة التسويقية للخدمات العامة؛ وأنها تروّج لسياسة توسعية وتزيد سلطة المتسلط؛ وأنها تفتقر إلى الاستدامة، والأدوات، وسهولة الاستخدام.

الخصخصة والهيكلة التسويقية للخدمات العامة:

برهن Jo Bates (2012) على أن «المبادرات المفتوحة مثل البيانات الحكومية المفتوحة (Open Government Data -OGD) تظهر كعملية تاريخية، وليست وقائع محايدة». فكما هو الحال مع كل المبادرات السياسية، فإن سياسة البيانات المفتوحة ليست ببساطة بديهية أو محايدة، بل هي ترتكز على عقيدة سياسية واقتصادية. فحركة البيانات المفتوحة متنوعة وتتكون من مجموعة من الدوائر الانتخابية ذات أجندات وأهداف مختلفة، وليس الدافع من ورائها مجرد أي طرف واحد منها. ومع ذلك، بينتBates بأن عامل الشدّ السياسي في حالة حركة البيانات المفتوحة في المملكة المتحدة كان ضئيلاً حتى بدأت الشركات الكبرى تنشط بتكتلها في حملات دعم البيانات المفتوحة، وبدأت المبادرات الحكومية المفتوحة تنسجم مع برامج التقشف القسري والهيكلة التسويقية للخدمات العامة. فبالنسبة لها،



قد استولت الأحزاب السياسية وقطاع الأعمال على حركة البيانات المفتوحة لصالح المصالح الرأسمالية المهيمنة تحت ستار أجندة / جدول أعمال الشفافية (Bates 2012).

وبعبارة أخرى، فإن جدول الأعمال الحقيقي لقطاع الأعمال هو الحصول على البيانات المنتجة بتكلفة باهظة دون أي مقابل، ومن ثم الوصول إلى البنية التحتية المدعومة بشكل كبير مما يمكنهم من الاستفادة من الأرباح، وفي الوقت ذاته، إزالة القطاع العام من السوق وإضعاف موقعه بصفته منتجاً لمثل هذه البيانات. ولذا، فبسبب اختفاء الدخل من خدمات البيانات (كما في الحالات التي جرى تمويلها من الرسوم وليس من دافعي الضرائب)، ومن ثم اختفاء التمويل اللازم لدعم الإنتاج والإدارة المحلية الداخلية، تم دفع هيئات القطاع العام إلى الاستعانة بمصادر خارجية لمثل هذه الخدمات من القطاع الخاص على أسس تنافسية أو التخلي عن إنتاج البيانات للقطاع الخاص التي عليهم بعد ذلك شراؤها منهم (Gurstein 2013). وهنا، يتعيّن شراء خدمات البيانات والبيانات المشتقة من البيانات المفتوحة المتاحة بحرية من منشئ البيانات، في الوقت نفسه الذي يتم إفراغ المعارف الأولية لبيانات المنظمة. وعلاوة على ذلك، وحيث إن البيانات المفتوحة تتعلق في كثير من الأحيان بأنشطة الهيئة نفسها، وخصوصاً عند اشتمالها على مؤشرات قياس الأداء الرئيسية، فهي تعمل على تسهيل إصلاح القطاع العام وإعادة تنظيمه مما يعزز الخصخصة، وروح الإدارة العامة الجديدة ومصالح القطاع الخاص (McClean 2011; Longo 2011). فمثل هذه العمليات، كما بيّنت Bates (2013)، هي جزء من إستراتيجية سياسية مرسومة لفتح وإتاحة كافة الخدمات العامة تقريباً للمنافسة بين مزودي الخدمات من القطاع الخاص والقطاعات الأخرى، حيث يعمل فتح البيانات الخاصة بالخدمات العامة على تمكين مستخدمي الخدمة من استحداث خيارات معلومة ضمن سوق الخدمات العامة على أساس التطبيقات التي تعتمد على البيانات التي ينتجها مجموعة من مطورو القطاعات التجارية وغير التجارية (2013، التعبير الأصلي). وفي مثل هذه الحالات، فإن جدول أعمال الشفافية التي يروّج لها السياسيون وقطاع الأعمال هو مجرد تعبير مجازي، فإذا كان أي من الطرفين مهتماً حقاً في الشفافية والحكومة المفتوحة فإنه سيكون داعماً بالتساوي للحق في حركة المعلومات والعمل المتعلق بإفشاء الأسرار (Janssen 2012) وكذلك تخفيف القيود عن حقوق الملكية الفكرية على نطاق أوسع (Shah 2013). وبدلاً من ذلك، فإن الحكومات والشركات بشكل عام مقاومة لهما على حد سواء.



السياسات التوسعية وتمكين المُمَكن:

هنــاك جــدل ذو صلة يتمثّــل في أن الغالب في حركــة البيانات المفتوحــة مدفوع من وجهة نظر فنية واقتصادية، ويركّز بشــكل كبير على إتاحة الوصول إلى البيانات، وليس على سياســات البيانات نفســها، وما قد تكشف عنه، أو كيفية اســتخدامها، ولمصالح من تصبّ (Shah 2013). وبعبارة أخرى، تســعى الحركة بشــكل كبير إلى تقديم صورة تظهرها على أنهــا محبّبة وبديهية، وتعزيــز الاعتقاد بأن فتح البيانات بطبيعتــه شيء جيد في حدّ ذاته. وبالنسبة لآخرين، فإن إتاحة الوصول إلى البيانات هو مجرد وجه واحد فيما يتعلق بمفهوم الانفتاح، وعلى القدر نفسه من الأهمية هو ما تتكوّن البيانات منه، والكيفية التي يمكن بها إنشاء مجتمع أكثر اعتدالاً وإنصافاً. فإذا كانت حركة البيانات المفتوحة تخدم فقط المصالح الرأســمالية من خلال فتــح البيانات العامة، وتحتفظ بالبيانــات الاحتكارية الخاصة خلف أنظمة جدران الدفع (بمقابل رســوم الاشــتراك) وحمايتها بأنظمــة الملكية الفكرية، وتمكن بصورة أكبر أولئك المُمَكنين بالفعل وحرمان الآخرين منها، فإنها تكون قد فشــلت في جعل المجتمع أكثر ديمقراطية وانفتاحاً (Gurstein 2011; Shah 2013).

ففكــرة أن البيانــات محايدة وموضوعية في طبيعتها تُطرح ضمناً في معظم المناقشــات المتعلقــة بالبيانــات المفتوحــة وأن للــكل إمكانية الوصول لهــذه البيانات واســتخدامها (Gurstein 2011; Johnson 2013). بيــد أن القضيــة ليســت كذلك، فمع الاحترام لفتح البيانات نفســها يؤكّد (2013) Johnson وجود درجة عالية من الامتياز الاجتماعي والقيم الاجتماعيــة الكامنــة ضمناً في بيانات القطاع العــام فيما يتعلق بماهيــة البيانات المنتجة، وتعلّقها بمن ولماذا (لا ســيما في المجالات التي تعمل كأنظمة تخصصية، مثل أنظمة الرعاية الاجتماعيــة، وأنظمة إنفاذ القانون)، التي تتمثّــل مصالحها ضمن مجموعة البيانات وأيضاً تلك التي يتمّ استبعاد مصالحها. وهكذا فإن هياكل القيمة متأصلة في مجموعات البيانات والتي تشــكّل بدورها التصميم والتفســير لها وربما تعمل على نشر المظالم وتعزيز المصالح المهيمنة.

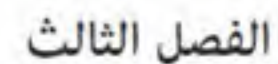
فالمواطنــون متفاوتون من حيث إمكانية الوصول إلى الأجهزة والبرامج المطلوبة لتحميل ومعالجة مجموعات البيانات المفتوحة، كما أن لهم مستويات مختلفة من المهارات اللازمة



لتحليل، وتأطير، وتفسير البيانات (Gurstein 2011). وحتى لو امتلكت بعض المجموعات القــدرة عــلى تقديم فهم مقنع للبيانــات، إلا أنهم لا يملكون بالــضرورة الاتصالات اللازمة لاكتســاب التأييد الجماهيري والتأثير على النقاش، أو الحنكة السياســية بمقابل خصم جيد التجهيز والدهاء. ومن ثم، فإن الديموقراطية المحتملة للبيانات المفتوحة مفرطة في التفاؤل، إذا أخذ بالحســبان معظم المستخدمين الذين لديهم درجة عالية من المعرفة التقنية وملف تعريفي سياسي قائم (McClean 2011)، فالواقع أنه يمكن للبيانات المفتوحة أن تعمل على زيادة تمكين السلطة وإعادة إنتاجها وتعميق اختلال توازن القوى (Gurstein 2011). ومن الأمثلة التي يستشهد كثيراً بها على هذا الأخير هو رقمنة سجلات الأراضي في ولاية كارناتاكا (Karnataka)، الهند، حيث عمل مشروع البيانات المفتوحة، والذي تمّ الترويج له باعتباره مبادرة لصالح الفقراء، على حرمان الفقراء من خلال تمكين ذوي الموارد المالية والمهارات من الوصول إلى البيانات التي ســبق حجبها وإعادة اقتطاع أراضيهم بوضع اليد بغية الاســتيلاء عليها (Gurstein 2011; Slee 2012; Donovan 2012). لقد ســهّلت البيانات المفتوحة في هذه الحالة، وبعيداً عن العمل على مساعدة كل المواطنين، على التغيير في حقوق ملكية الأرض ونقل الثروة من الفقراء إلى الأغنياء. وبعبارة أخرى، فإن فتح البيانات لا يعني عملية متأصلــة في الديمقراطية وتعزيزها، بل إن الواقع أنه يُمكن للبيانات المفتوحة أن تعمل أداة للسلطة التأديبية (Johnson 2013).

ومــن المهــم هنا أن التنبّــه إلى أن إنتاج البيانــات الحكومية يكون لأغــراض الحوكمة، وهي تتألف من نوعين رئيســيين: الأنشــطة المتعلقة بأنشــطة الدولة والأنشــطة المتعلقة بالمواطنــين، والأماكن، والأعــمال. فالنوع الأول يهتمّ بكيفية عمل الدولة، وعند إتاحة هذه البيانات بشــفافية، يمكن اســتخدامها لتقييم الأداء والمســاءلة. وقد يقــف قلة ضدّ جعل مثل هذه البيانات متاحة للجمهور، على الرغم من الحاجة لتقدير مثل سياســات المكاشفة هذه، خاصة إذا كانت التدابير المســتخدمة ذات تأثير مشــوّه لتقديم الخدمة وتعمل على التلاعــب ببيانــات التقييم (بالتوافق مع قانون كامبــل -Campbell's Law- انظر الفصل الســابع)، وهذا يؤثر بدوره ســلباً في ماهية الخدمات المقدمة وجعل الحياة ســيئة لأولئك الأشــخاص الذين يتلقون الخدمة. ويتكوّن النوع الثاني من السجلات الشخصية والمؤسسية الحساســة للغاية، ولم يتم إنشــاؤها بهدف مشــاركتها وتبادلها مع الآخريــن. وفي الواقع،





يتوقـع المواطنون لها أن تكون محمية بموجب قوانين الخصوصية وحماية البيانات. وحتى عندمـا يتـم تجريد هذه البيانات من البيانات التي تشـير إلى هوية أشـخاص معينين أي جعلهـا مجهولـة (Anonymised) وعرضها مجملـة، فإنه يمكن للبيانـات أن تكون على درجة من الحساسـية والسياسـية بصورة كبيرة. فخذ، على سـبيل المثـال، بيانات الرعاية الاجتماعية والصحية المنقحة نسبياً والمجمعة في وحدات مكانية (على سبيل المثال، مستوى الحـي)، فهذه البيانات تُمثل أداة لتوجيـه التدخلات الموجهة الرامية إلى معالجة المعوقات الاجتماعية، كما أن لها إسـهامات مفيدة في تحليل البيانات التي تسـعي اجتماعياً إلى فرز المواطنين والتعريف عنهم فيما يتعلق بالقروض الائتمانية والتأمين ضد المخاطر، ويمكن من ثم اسـتخدامها لإنشاء ملفات تعريف مناطقية والتي توصم موضعاً ما وتحدّ من الاستثمار فيـه (انظر الفصل العـاشر). وبعبارة أخرى، يمكن إعادة توظيف البيانات لأغراض مختلفة بطرق شتى والتي لها مخرجات متباينة، وهناك أسباب مشروعة لتوخي الحذر فيما يتعلق بما يتم الإفراج عنه من البيانات الحكومية ومقاومة شعار التبسيط المستخدم من قبل بعض المدافعـين عن البيانات المفتوحة "إنها بياناتنا، لقد دفعنا مقابلها، وينبغي أن يكون لنا حق الوصول إليها".

الاستدامة والفائدة وسهولة الاستخدام:

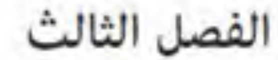
تركّـز الاهتمام بشـكل كبير، حتى الآن، عـلى جانب توفير الوصول إلى البيانات وإنشـاء مبادرات البيانات المفتوحة، وكان هناك عدم إيلاء اهتمام كاف لاسـتدامة المبادرات والكيفية التي يتم بها استخدام البيانات وتوظيفها. ففي دراسة أجريت على عدد من مشاريع البيانات المفتوحة المختلفة أوضحت (2012) .Helbig et al بأن العديد منها تركز بشدة على الناحية التقنية الفنية، ويمكن فهم هذه المشـاريع عـلى أنها «تزيد قليلاً عن المواقع المرتبطة بملفات بيانات متنوعة، ودون أي مراعاة لسهولة الاستخدام، وجودة المحتوى، أو عواقب استخدامها». والنتيجـة هـي مجموعة من مواقع البيانـات المفتوحة التي تعمل أكـثر كحاويات للبيانات أو مكبـات البيانات، بحيـث تفتقر إلى الصفات المتوقعة في البنيـة التحتية العاملة للبيانات والمنظمـة تنظيماً جيداً، مثل البيانات النظيفة، ذات الجودة العالية، والتي جرى التحقق من صحتهـا، وقابليتها للتشـغيل المتبادل بما يتوافق مع معايير البيانـات ولها التعريف الوصفي



المناسب ومجموعات تسجيل كاملة (الوثائق المرتبطة بها)؛ أي سياسات المحافظة والنسخ الاحتياطي والتدقيق؛ وسياسات إعادة الاستخدام والخصوصية والأخلاقيات، والترتيبات الإدارية وآليات تنظيم الإدارة والحوكمة؛ والاستقرار المالي وخطة طويلة الأجل للتنمية والاستدامة (انظر الفصل الثاني). كما تفتقد العديد من المواقع للأدوات المناسبة والمواد السياقية لدعم تحليل البيانات، وعلاوة على ذلك فإن غالب مجموعات البيانات المنشورة تكون صعبة الأخذ والاستفادة منها، حيث تتألف من تلك التي يسهل نشرها وتحتوي على بيانات غير حساسة كما أن لديها فائدة منخفضة نسبياً. وفي المقابل، غالباً ما تبقى البيانات الأكثر صعوبة وتطلّباً لجعلها مفتوحة في حالة مغلقة، بسبب قضايا حساسيتها أو لأنها تتطلب المزيد من العمل الإداري للامتثال لقوانين حماية البيانات (Chignard 2013).

ويتمثل جزء من المشكلة في أن العديد من مواقع البيانات المفتوحة هي استجابة عنيفة وجاهزة لظاهرة ناشئة، وقد تمّ بناؤها من قبل متحمسين ومنظمات ممن لديهم خبرة قليلة في أرشفة البيانات أو الاستخدام السياقي للبيانات التي تمّ فتحها. وقد تمّ دعمها والترويج لها من خلال الملتقيات المكثّفة لمطوري البرمجيات (الهاكاثونات، Hackathons) ومنتديات العطل الأسبوعية لمجموعات الاختصاص (Data Dives)، التي تعيد طرح وإنتاج العديد من هذه القضايا. وكما أكّد McKeon (2013) and Porway (2013) فإن هذه الأحداث والتجمعات التي تدعو المبرمجون والأطراف المعنية الأخرى لبناء تطبيقات باستخدام البيانات المفتوحة لها فائدة محدودة. ففي حين أنها تركّز الاهتمام على البيانات وتُعد وسيلة جيدة للتواصل، إلا أن أولئك الذين يقومون بالبرمجة غالباً ما يفتقرون إلى المعرفة السياقية المعمّقة فيما يتعلق بما تشير البيانات إليه، أو ينتمون إلى مجموعة سكانية معينة ليست بالضرورة معبّرة عن شريحة مجتمعية أوسع نطاقاً (على سبيل المثال، الشباب والمتعلمون والتقنيون)، ويظنّون أن المشاكل الهيكلية العميقة يمكن حلّها من خلال الحلول التقنية. إنهم «كيان مؤلف من مجتمع صغير من المتطوعين المتحمسين، وليسوا من الأشخاص الذين لديهم مصلحة كبيرة في رؤية المشروع ينجح» (McKeon 2013). وعلاوة على ذلك، لا تزال الحلول التي تمّ إنشاؤها في هذه اللقاءات على إصدارها الأول، مع قليل من المتابعة بعد انتهاء الحدث ودون تحديث أو صيانة أو تطوير لاحق. وكان جواب Porway's (2013) لبناء حلول تشرّع لتغيير مناسب هو الجمع بين أولئك الذين يفهمون





البيانات والمشكلات القائمة على أرض الواقع مع أولئك الذين يستطيعون برمجة التطبيقات وبنائها والمحافظة عليها، وقد اقترح McKeon (2013) عقد اجتماعات تخطيطية مسبقة وتوصيات توجيهية بعد الحدث تهدف إلى استدامة المخرجات وإنشاء مجتمع على المدى الطويل. وبعبارة أخرى، هناك حاجة للتوفيق بين الخبرات الموجودة والخبرات في تطوير التطبيقات. والحاجة نفسها مطلوبة مع مواقع البيانات المفتوحة، حيث تتطلب تعلّم الدروس المستفادة من أولئك الذين عملوا لفترة طويلة سابقة في الأرشفة وقطاع أعمال تبادل البيانات، متضمناً ذلك أمناء المحفوظات الأرشيفية والعلماء والجغرافيون.

ونتيجة لهذه القضايا العويصة المختلفة، وبدلاً من إنشاء دورة إيجابية (Virtuous Cycle) بحسب ما تفترض حركة البيانات المفتوحة، حيث الإفراج عن المزيد والمزيد من مجموعات البيانات، في المزيد من الأشكال والصيغ، ينتج عنه مزيد من الاستخدام، ومن ثم الإفراج عن المزيد من البيانات مستقبلاً، إلا أن العديد من المواقع، كما لاحظت (2012) .Helbig et al عانت من انخفاض معدلات التصفح لها وتراجعها لأنها لا تشجع على الاستخدام أو لا تقدم التسهيلات للمستخدمين في هذا الخصوص، وتقتصر خدماتها على عوامل أخرى مثل ممارسات إدارة البيانات وعرض الجهد المؤسسي المبذول والسياسات الداخلية المعمول بها. فبعد الشرارة الأولى من الاهتمام، تراجع استخدام البيانات بشكل ملحوظ للغاية بسبب الكشف عن القيود المفروضة على البيانات ومحدوديتها ونضال المستخدمين من أجل إيجاد طرق لكيفية استخدام وتحليل البيانات بشكل مربح. وقد لاحظ McClean (2011)، على سبيل المثال، أن التحليل الناشئ عن البيانات المفتوحة كان له تأثير محدود على النقاشات السياسية، وخلص فيما يتعلق بنظام المعلومات المجمّعة على الإنترنت (كوينز COINS) الخاص بالبيانات المالية الحكومية في المملكة المتحدة، إلى أنه بعد

هبّة قصيرة من اهتمام وسائل الإعلام في منتصف عام 2010، وفي أعقاب إطلاقه.... فإن التقارير التي تشير صراحة إلى نظام كوينز غدت الآن نادرة للغاية، كما أن العاملين في الصحافة الذين كانوا الأكثر اهتمامًا بالحصول على حق الوصول إليه قد صرّحوا بأنه لم تثبت فائدته كمحرّك فاعل للصحافة على وجه الخصوص.

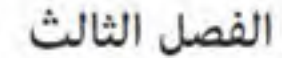


وحيــث يتم نشر البيانات دورياً على ســبيل المثــال، بصورة فصلية أو ســنوية، فإن الاستخدام يميل إلى أن يكون دورياً وغالباً مرتبطاً بمشاريع محددة مثل التقارير الاستشارية بدلاً من وجود نمط أكثر ثباتاً من الاستخدام الذي قد يتوقعه المرء لبيانات تتغير بطبيعتها آنياً بصورة مســتمرة أو لديها قيمة اقتصادية عاليــة. ففي مثل هذه الحالات، لاحظت Helbig et al. (2012) أن مجموعة من حلقات ردود الفعل الســلبية قد أبطأت توفير البيانات واســتخدامها، ومن ثم زيادة خفض الاســتخدام. ولذا، فبعد بعض «المكاســب السريعة» الأولية، يكمن الخطر في تحوّل أي دورة إيجابية من كونها إيجابية إلى سلبية، وتقويض الأســاس المنطقي لتمويل الحكومــة المركزية لمثل هذه المبادرات وكونها عرضة للتوقــف المفاجئ. إن المفتاح لتجنب نشــوء مثل هذه الدورة الســلبية هو التأكد من تركيــز أي مبادرة بأكبر قدر ممكن على جانــب الطلب كما في جانب العرض، وأنها توفّر للمســتخدمين بيانات قابلة للتشــغيل المتبادل والأدوات التحليلية وغيرها من الخدمات التي تسهل استخدام وإضافة قيمة للبيانات، بدلاً من مجرد ربط الملفات.

الخلاصة:

عند أحد المستويات، تُعد قضية البيانات المفتوحة والبيانات المرتبطة بديهية - حيث تُسهم البيانات المفتوحة في صنع الشفافية والمساءلة والمشاركة والاختيار والابتكار الاجتماعي والكفاءة والإنتاجية وتعزيز الحوكمة والابتكار الاقتصادي وتكوين الثروات. في حين تعمل البيانات المرتبطة على تحويل المعلومات الموزعة عبر شبكة الإنترنت إلى شبكة دلالية يمكن من خلالها القراءة الآلية للبيانــات وربطها معــاً. وعليه تحمل البيانات المفتوحة والبيانات المرتبطة في طياتها الكثير من الوعود والآمال والقيمة مع اعتبارها مجازفة محفوفة بالمخاطر، فواقع الحال بالنســبة للبيانات المفتوحــة والبيانــات المرتبطة أكثر تعقيداً، كما أن الأســس الاقتصادية لها ليســت واضحة على الإطلاق. ويبدو أن للبيانات المفتوحة والبيانات المرتبطة تكاليف هامشــية قليلة، ولكن الحاجة قائمة في عملية إنتاجها والطاقم الفني والجهاز المؤسسي لتسهيل استخدامها والحفاظ عليها مما يلــزم تكاليف حقيقية من حيث العمالة، والمعدات، والموارد. وفي الواقع، فإن البيانات المفتوحة بعيدة كل البعد عن كونها مجانية، ولكن حتى الآن، لم يتمّ وضع نموذج متماسك لتمويل مستدام





لمبادرات البيانات المفتوحة دون إعانة حكومية. كما أن فوائد البيانات المفتوحة من حيث إنتاج قيمة مضافة على فائض الاستهلاك والمشاريع التجارية الجديدة، والمنتجات المبتكرة، والتكاليف قد قابلها إيرادات ضريبية إضافية افتراضية إلى حد كبير أو بدأت بالتحقق فقط، وعلاوة على ذلك، فإن العواقب السلبية المحتملة لفتح البيانات لم تستكشف بعد بالكامل.

فهناك حاجة إلى الدفع بمزيد من الاهتمام النقدي لكيفية تطور مشاريع البيانات المفتوحة كظاهرة تقنية اجتماعية مقعدة ومرتبطة بطائفة متنوعة من البرامج، والمعنيين، وأصحاب المصلحة. وحتى الآن، تمحورت الجهود حول العمل السياسي والتقني المتعلق بإقامة مشاريع البيانات المفتوحة، ولم يتم العمل بشكل كاف على دراسة هذه التحركات المادية والتي تلامس العديد من الجوانب وعواقبها. ونتيجة لذلك، فإننا نفتقر إلى دراسات الحالة التفصيلية لتطبيق مشاريع البيانات المفتوحة، والتجمعات المحيطة بها، والتي تعمل على صياغتها، والطرق الفوضوية، والطارئة والعلائقية التي تتكشف بمرور الوقت. ولن تظهر صورة أكثر اكتمالاً للبيانات المفتوحة والبيانات المرتبطة إلا من خلال مثل هذه الدراسات التي ستكشف عن كل من الإيجابيات والسلبيات لمثل هذه المشاريع وتوفر إجابات عن أسئلة معيارية بشأن الكيفية التي ينبغي تنفيذها بها ومعالجة قضايا الاستدامة وسهولة الاستخدام والسياسات المتعلقة بها.

الفصل الرابع

البيانات الكبيرة (Big Data)

تتبّع (2012) Francis Diebold منشأ استخدام مصطلح «البيانات الكبيرة» إلى منتصف التسعينيات (1990s)، حيث استخدم لأول مرة من قبل جون ماشي (John Mashey)، كبير العلماء السابق المتقاعد من شركة سيليكون غرافيكس، للإشارة إلى معالجة وتحليل مجموعات البيانات الضخمة. وقد كان المصطلح قليل الجذب في البداية. ففي العام ٢٠٠٨ استخدم عدد قليل جداً من الناس مصطلح «البيانات الكبيرة»، سواء في المجالات الأكاديمية أو الصناعية. وبعد خمس سنوات أصبح مصطلح «البيانات الكبيرة» التعبير الطنّان الأكثر شيوعاً في الاستخدام في دوائر الأعمال ووسائل الإعلام الشعبية، مع تعليقات دورية في الصحف الأكثر انتشاراً، مثل صحيفة النيويورك تايمز (New York Times) والفايننشال تايمز (Financial Times)، ومقاطع وأعداد خاصة في المجلات العامة والمجلات العلمية مثل عالم الاقتصاد (Economist) والوقت (Time) والطبيعة (Nature)، والعلوم (Science). لقد كان هذا الانتشار وزيادة الاستخدام المرتبطة به هو ما أدّى بالفعل إلى إعلان غارتنر (Gartner) في يناير (2013) عن انتقال تقنية البيانات الكبيرة على طول دورة الدعاية (Hype cycle) (كما يطلق عليها نموذج النضج التقني) من مرحلة «ذروة تضخم التوقعات» إلى «مستوى خيبة الأمل» (Sicular 2013)، مع إعلان بعض الدعاة سابقاً عن موت «البيانات الكبيرة» بصفته مصطلحاً ذا معنى، حيث أصبح مصطلحاً متشعباً وواسع النطاق للغاية في التعريف (على سبيل المثال، de Goes 2013)، فبعض أوائل الشركات التي تبنّت المفهوم تصارع من أجل تحويل الاستثمارات في البيانات الكبيرة إلى عوائد ربحية والبعض يشككون في فوائدها المحتملة. ومع ذلك، ظلّ قطاع الأعمال والقطاع الحكومي العام وممولو البحوث على قناعتهم بارتفاع منحنى التنوير (Slope of Enlightenment) للبيانات الكبيرة مجدداً إلى مستوى الإنتاجية (Plateau of Productivity) عبر دورة الإعلان أو نموذج النضج التقني، بل وأكثر من ذلك، أنها وُجدت لإحداث تغيير جذري في كيفية عمل العلوم وإجراء الأعمال (Siccular 2013، انظر أيضاً للفصلين السابع والثامن).



وكما نوقش في الفصل الثاني، فقد وُجدت ولفترة طويلة من الزمن مجموعات ضخمة من بيانات عبر الحكومات والصناعات والأوساط الأكاديمية التي تمّ استخراج المعلومات منها من أجل تقديم مزيد من الرؤى والمعرفة، فالإدارات والوكالات الحكومية أنتجت بشكل روتيني كميات ضخمة من البيانات. على سبيل المثال، قامت إدارة المحفوظات والسجلات الوطنية في الولايات المتحدة في عام ٢٠١٣ بتخزين ما يقارب ٤,٥ مليون قدم مكعب من الوثائق المادية من وكالات السلطة التنفيذية الأمريكية، والمحاكم، ومجلس النواب، ومقر الرئاسة (وهي خمسة في المائة من سجلات الحكومة الاتحادية) مما يضيف ثلاثين ألف قدم بصورة خطية سنوياً (Ellis 2013)، بالإضافة إلى خمسمائة تيرابايت من البيانات الرقمية. كما تمّ بالمثل جمع بيانات للشركات عن عملياتها، والأسواق، والعملاء وقواعد بيانات هائلة من البيانات العلمية التي تمّ جمعها ومشاركتها من بداية العلم الحديث. لذلك، ما المقصود من مصطلح «البيانات الكبيرة»، بالنظر إلى كميات البيانات الناتجة في المجالات السابقة؟

فكما هو شأن العديد من المصطلحات المستخدمة للإشارة إلى الاستخدام المتسارع في التقنيات والممارسات، لا يوجد تعريف أكاديمي محدد أو صناعي متفق عليه لمصطلح البيانات الكبيرة، إلا أن الكثير من التعاريف الشائعة تشير إلى الثلاثية المكررة للأحرف الابتدائية (3Vs) للخصائص المميزة للبيانات الكبيرة: الحجم (Volume)، والسرعة (Velocity)، والتنوع (Variety) (Laney 2001; Zikopoulos et al. 2012). فالبيانات الكبيرة تتميز بأنها:

- ضخمة في الحجم، أي تتكوّن من تيرابايت أوبيتابايت من البيانات.
- عالية السرعة، أي أنها تنشأ بشكل آني أو فوري، أو قريباً من الزمن الحقيقي.
- مختلفة من حيث تعدد تنوّعها، حيث تتنوّع البيانات بين المنظمة وغير المنظمة في طبيعتها، كما تتنوع في الغالب من حيث مرجعيتها الزمانية والمكانية.

فخلال زمن ما قبل البيانات الكبيرة، تقيّدت قواعد البيانات بين هذه الخصائص الثلاث: حيث كان من الممكن لقاعدة البيانات الجمع بين خاصيتين اثنتين في آن واحد (ضخمة



وسريعة، متنوعة وسريعة، وكبيرة ومتنوعة) (Croll 2012). إلا أنه مع تعزيز قدرة المعالجة الحاسوبية، والتصاميم الجديدة لقواعد البيانات، والتخزين الموزّع (انظر الفصل الخامس)، أصبح تحقيق الثلاثة مجتمعة في وقت واحد ممكناً مما يتيح أشكالاً جديدة للتحليل وتقديم عروض مفصلة للغاية للنظم الضخمة والمتغيرة باستمرار. وعدا عن الخصائص الثلاث السابقة، أشارت المؤلفات الصادرة مؤخراً إلى عدد من الخصائص الرئيسية الأخرى، حيث تتميز البيانات الكبيرة أيضاً بالميزات التالية:

- شاملة في نطاقها، حيث تسعى لتغطية كامل مجتمع أو نظم الدراسة (ن = الجميع)، أو على الأقل توفر حجم عينات أكبر من تلك التي يجري استخدامها في دراسات البيانات التقليدية أو الصغيرة.
- متناهية في دقتها، حيث تهدف إلى أن تكون مفصّلة قدر الإمكان، ومفهرسة بشكل فريد في تحديد الهوية.
- علائقية بطبيعتها، كونها تحتوي على حقول بيانات مشتركة تمكّن من الربط بين مجموعات بيانات مختلفة.
- مرنة، حيث تحمل سمات التفرّع (يمكن إضافة حقول بيانات جديدة بسهولة) مع قابلية التوسع (يمكن التوسع في حجمها بسرعة).

boyd and Crawford 2012; Dodge and Kitchin 2005; Marz and Warren)
.(2012; Mayer- Schonberger and Cukier 2013

وبالنظر إلى الحملة الدافعة إلى التحويل الرقمي (الرقمنة- Digitisation) للبيانات التقليدية الصغيرة والتوسع فيها ضمن البنى التحتية الرقمية التي تتميز بالضخامة والتنوع مثل المحفوظات الأرشيفية الوطنية، التعدادات السكانية، والمجموعات الثقافية والاجتماعية التراثية؛ انظر الفصل الثاني)، فقد كانت السرعة والخصائص الإضافية هي ما ميزت البيانات الكبيرة وجعلت منها ابتكاراً استثنائياً (Christensen's 1997) وغيّر بصورة جذرية طبيعة البيانات وما يمكن عمله من خلالها (انظر الجدول ١-٢). فعلى سبيل المثال، يتميز المسح الأسري الوطني بكبر الحجم، والدقة المتناهية، وقوة العلائقية، إلا أنه يفتقر إلى السرعة

(تتغير البيانات مرة واحدة في السـنة)، والتنوع (يتكوّن من ثلاثين سؤالاً منظماً)، والشمول (حيـث تقتـصر العينة على فرد واحد مـن عشرين أسرة)، والمرونة (كـون حقول البيانات ثابتة ولا تتغير خلال المسـوحات المتعاقبة للتمكن من تحليل السلاسل الزمنية). ففي هذا الفصل تمّ إيراد الخصائص السبعة للبيانات الكبيرة ويناقش الفصل التالي العوامل المساعدة للبيانات الكبيرة ومصادرها.

الحجم (Volume):

شـهد العقد الماضي انفجاراً في كمية البيانات التي تمّ إنتاجها ومعالجتها بشـكل يومي. وكما عنونت مجلة وايرد (Wired) بشكل رئيس في عددها الخاص للعام 2008: نحن ندخل «عـصر التيرابايت» (وفي الواقع، لقد ولجنا بالفعل في عصر الزيتابايت؛ 2^70 بايت). وقد سعت العديد من الدراسات لتقدير وتتبّع الأحجام المعنية (على سبيل المثال، Hilbert and López 2009; Gantz and Reinsel 2011; Short et al. 2011). واسـتخدمت هـذه الدراسات منهجيات وتعاريف مختلفة، ولكنها أجمعت بالكلية على أن معدل نمو البيانات كان رهيبـاً في نطاقـه، وعـلاوة على ذلك، فهي قائمـة على النمو باطراد أكبر في المسـتقبل المنظور. إن أبسط طريقة لتوضيح هذا النمو هو إعطاء بعض الأمثلة على تقديرات عالمية لأحجام البيانات وبعض التقديرات المتعلقة بكيانات محددة، ولتوفير إطار مرجعي، يفصّل الجدول (١-٤) بصورة مختصرة كيفية قياس حجم البيانات.

الجدول (١-٤)

مقاييس البيانات الرقمية

الوحدة	القياس	ما الذي تعنيه
بت (b)	صفر أو1	هي اختصار لتعبير «الخانة الثنائية»، حيث تستخدمه الحاسبات لتخزين ومعالجة البيانات على صورة الرمز الثنائي (0 أو1)
بايت (B)	ثمانية خانات ثنائية (8 بت)	مساحة تخزينية كافية لتمثيل رموز الأبجدية الإنجليزية والأرقام باستخدام الرموز الثنائية



كيلوبايت (KB)	1000 أو210 بايت	من كلمة «الألف» (Thousand) باليونانية. فالصفحة الواحد من النص المطبوع تقدّر بحوالي 2 كيلوبايت
ميجا بايت (MB)	1000 كيلوبايت؛ 220 بايت	من كلمة «ضخم» (Large) باليونانية. فكامل أعمال شكسبير الأدبية يقدّر حجمها ب 5 ميجابايتات في حين يقدّر حجم الملف الصوتي الموسيقي بحوالي 4 ميجابايت
غيغابايت (GB)	1000 ميجابايت؛ 230 بايت	من كلمة «هائل» (Giant) باليونانية. ففيلم من ساعتين يمكن ضغطه في ملف حجمه 1-2 غيغابايت
تيرابايت (TB)	1000 غيغابايت؛ 240 بايت	من كلمة 'عملاق' (Monster) باليونانية. فالكتب المفهرسة في مكتبة مجلس النواب الأمريكي (الكونغرس) يقدّر حجمها بحوالي 15 تيرابايت
بيتابايت (PB)	1000 تيرابايت؛ 250 بايت	تقدّر بيانات كل الرسائل التي تمّ توصيلها من قبل خدمة البريد الأمريكي في العام 2010 بنحو 5 بيتابايت.
إكسابايت (EB)	1000 بيتابايت؛ 260 بايت	مكافئة لبيانات عشرة بلايين نسخة من مجلة عالم الاقتصاد (The Economist).
زيتابايت (ZB)	1000 إكسابايت؛ 270 بايت	يقدّر المجموع الكلي للمعلومات في الوجود في عام 2010 بحوالي 1.2 زيتابايت
يوتابايت (YB)	1000 زيتابايت؛ 280 بايت	حالياً هي قيمة أكبر من التخيّل
تمّ وضع الاختصارات من قبل المنظمة الحكومية الدولية intergovernmental organization، والمكتب الدولي للأوزان والمقاييس International Bureau of Weights and Measures. تمّ إضافة وحدتي يوتابايت وزيتابايت في العام 1991؛ المصطلحات للكميات الأكبر لم يتمّ وضعها بعد.		

المصـــدر: مجلــة عالم الاقتصــاد (2010) (The Economist)، تمّ إعــادة إنتاجها بإذن عالم الاقتصــاد. صحيفة عالم الاقتصاد المحدودة، لندن، الإصدار 11 مارس 2014.

وقد بيّن (2012) .Zikopoulos et al أنه في العام ٢٠٠٠م تمّ تخزين ٨٠٠٠٠٠ بيتابايت مــن البيانات في العالم. وبحســب التقريــر الســنوي ل (2011: 7) .Short et al –كَمْ كَمّ المعلومات؟ - عالجت خادمات العالم للعام ٢٠٠٨م، 9.57 زيتابايت من المعلومات... وكان هــذا بمعدل ١٢ غيغابايــت من المعلومات يومياً للعامل العــادي، أو نحو ٣ تيرابايتات من



المعلومات لكل عامل سنوياً. وقامت الشركات في العالم بمعالجة ٦٣ تيرابايتاً من المعلومات سنوياً باستثناء المصادر غير المحوسبة، وبحلول عام ٢٠١٠ م، قدّرت مؤسسة إم جي آي (MGI) (بحسب ما ذكر Manyika et al. 2011: 3) بأن الشركات عالمياً قد قامت بتخزين أكثر من ٧ إكسابايت من البيانات الجديدة على الأقراص الصلبة ... في حين قام المستهلكون بتخزين أكثر من ٦ إكسابايتات من البيانات الجديدة على أجهزة من قبيل أجهزة الحاسب الشخصية وأجهزة الحاسب المحمولة. كما قدّروا كذلك أنه في «عام ٢٠٠٩ م، كان ما يقرب من كلّ القطاعات في الاقتصاد الأمريكي لا يقلّ بياناتها المخزّنة بالمتوسط عن ٢٠٠ تيرابايت من البيانات لكل شركة فيها أكثر من ألف موظف. كما امتلكت العديد من القطاعات أكثر من 1 بيتابايت من البيانات المخزّنة لكل شركة». وفي عام ٢٠١٣ م، أشار Neelie Kroes، مفوّض الاتحاد الأوروبي لجدول الأعمال الرقمي، بأن ١,٧ مليون بليون بايت من البيانات يجري توليدها في الدقيقة الواحدة على مستوى العالم (Rial 2013).

وبناء على مراجعة Manyika et al. (2011) لنمو حجم البيانات، فقد توقعوا ارتفاعاً بنسبة ٤٠ في المائة في نمو البيانات التي يتمّ إنشاؤها على مستوى العالم سنوياً. وقدّر Gantz and Reinsel (2011) بأن «كمية المعلومات التي تمّ إنشاؤها وإعادة نسخها (تكرارها) على الإنترنت ستتجاوز ١,٨ زيتابايت (١,٨ تريليون غيغابايت) في العام ٢٠١١ مخزّنة في ٥٠٠ كوادريليون (مليون مليار) ملف بيانات». ويبيّن هذا الذي ذكروه نمواً بمقدار تسعة أضعاف خلال خمس سنوات فقط، مع نمو متوقع في ذات الوقت إلى أكثر من الضعف كل سنتين. ونتيجة لذلك، فقد توقعوا في تقريرهم بأنه في العقد التالي:

سيزداد عدد الخوادم (الافتراضية والفيزيائية) في جميع أنحاء العالم بعشرة أضعاف، في حين ستزداد كمية المعلومات المُدارة في مراكز بيانات الشركات بمقدار خمسين ضعفاً، وسيزداد أعداد الملفات التي سوف تضطر مراكز البيانات إلى التعامل معها بمقدار 75 ضعفاً، على الأقل.

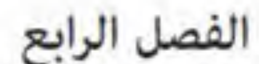
وبمثل هذا النمو الهائل في إنتاج البيانات، ادعت شركة آي بي أم IBM (2012) بأن «٩٠٪ من البيانات في العالم اليوم قد تمّ إنشاؤها في العامين الماضيين فقط»



كـما توقّع (2012) .Zikopoulos et al أن تقارب أحجام البيانات حدّ ٣٥ زيتابايت بحلول عام ٢٠٢٠م.

ويمكننـا إدراك وقـوع هذا الانفجار في إنتاج البيانات الرقميـة من خلال النظر في بعض الأمثلـة المحددة. فشركة تـك أمريكا (TechAmerica) قدّرت بأنه يجري يومياً إرسـال ما يقرب من 114 مليار رسـالة بريد إلكترونية و٢٤ مليار رسالة نصية، كما يتم إجراء ١٢ مليار مكالمة هاتفية على مسـتوى العالم (Strohm and Homan 2013). ووفقاً لشركة سيسكو (CISCO)، ففي العام ٢٠١٣ م كان هنالك تقريباً ما يناهز ١٠ مليارات عنصر (أجهزة اتصال ومستشـعرات) مشكّلة شبكة إنترنت الأشياء (Internet of Things -IoT)، والتي يقوم كلّ منهـا بإنتاج كميـات متغيرة من البيانات، مع الزيادة المتوقعة لهذا الرقم ليفوق الخمسـين مليـاراً بحلول عـام ٢٠٢٠ م (Farber 2013). أما ما يتعلّق بحركـة الاتصال على الإنترنت، فقـد كانـت شركة جوجل تعالج ٣ مليارات طلب اسـتعلام للبحث يوميـاً، وتقوم بتخزينها جميعـاً (Mayer-Schonberger and Cukier 2013)، وقرابـة ٢٤ بيتابايـت من البيانات كل يـوم (Davenport et al. 2012). وفي العـام ٢٠١١ م، قضى المسـتخدمون النشـطون لموقع التواصل الاجتماعي فيسـبوك أكثر من ٩,٣ مليارات سـاعة خلال شهر (Manyika et al. 2011)، وبحلـول عام ٢٠١٢ م، ذكـرت الشركة بأنها كانت تعالج ٢,٥ مليار من مختلف أنـواع بيانات المحتـوى (وصلات، مخازن، صور، أخبار، وما إلى ذلـك)، و٢,٧ مليار من ردود الإعجـاب وثلاثمائة مليون عمليـة تحميل للصور يوميـاً (Constine 2012). كما تمّ إنتاج أكـثر من أربعمائة مليـون تغريدة يوميـاً في العام ٢٠١٢م، بمعدل زيـادة مقداره مئتان في المائة سـنوياً، مع احتـواء كل تغريدة على ثلاثـة وثلاثين بنداً مـن البيانات الوصفية لهـا (Mayer-Schonberger and Cukier 2013). وتُعـد الكثير من هذه البيانات غير منظمة بطبيعتها. وبالمثل هنالك انفجار في حجم إنتاج البيانات المنظمة قد وقع أيضاً، على سبيل المثال، وفيما يتعلق ببيانات تجارة التجزئة الخاصة بالأوراق المالية والمبيعات، والتي تمّ جمعها من خلال السلاسـل السوقية (اللوجستية) ومنافذ البيع، أنتجت سلسلة متاجر وول مارت (Walmart) أكثر من ٢,٥ بيتابايت من البيانات المتعلقة بأكثر من مليون عملية خاصة بالعملاء في كل سـاعة في العام ٢٠١٢م، «أي يفوق ما يعادل ١٦٧ مرة من المعلومات الواردة في جميع الكتب المصنفة في مكتبة الكونغرس»؛ (Open Data Center Alliance 2012: 6)،





كما أنتجت متاجر تسكو (Tesco) البريطانية أكثر من ١,٥ مليار بنداً جديداً من بنود البيانات كل شهر في العام ٢٠١١ م (Manyika et al. 2011).

وبالمثل، تنتج الحكومات والهيئات العامة كميات هائلة من البيانات حول مواطنيها والدول الأخرى. على سبيل المثال، بدأت هيئات التنقل والعبور في مراقبة التدفق المستمر للناس من خلال أنظمة النقل، على سبيل المثال، جمع زمن ومكان استخدام بطاقات السفر مسبقة الدفع مثل بطاقة أويستر في لندن. كما يجري الآن تقديم العديد من نماذج وطرق دفع الضرائب أو طلبات الحصول على الخدمات الحكومية من خلال الانترنت، ففي العام 2009م، أنتجت حكومة الولايات المتحدة ٨٤٨ بيتابايت من البيانات (مؤسسة التقنية الأمريكية TechAmerica Foundation 2012). كما تعمل وكالات الاستخبارات الست عشرة التي تشكل أمن الولايات المتحدة، جنباً إلى جنب مع فروع الجيش الأمريكي، على مراقبة وتخزين وتحليل كميات هائلة من البيانات كل ساعة، من خلال آلاف المحللين العاملين في تدقيق وتفسير النتائج. وللحصول على فكرة عن حجم بعض مشاريع الاستخبارات العسكرية، فإن مشروع النظام المستقل للمراقبة الأرضية الصورية الحقيقية (ARGUS-IS)، والذي كشفت عنه وكالة مشاريع البحوث المتطورة الدفاعية (داربا DARPA) وجيش الولايات المتحدة في العام ٢٠١٣ م، هو منصة للمراقبة بالفيديو بدقة ١,٨ غيغابايت والتي يمكنها بيان التفاصيل الصغيرة حتى ست بوصات من ارتفاع ٢٠٠٠٠ قدم (ستة كيلومترات) (Anthony 2013). وتجمع المنصة ١,٨ مليار بكسل، في ١٢ إطاراً في الثانية { لقطة في الثانية }، مما ينتج ما يقارب ٦٠٠ غيغابت في الثانية. وهذا يعادل نحو ٦ بيتابايتات من البيانات يومياً. وباستخدام الحاسوب العملاق، يمكن للنظام إجراء التحليل آنياً من خلال تتبّع ما يصل إلى ٦٥ جسماً متحركاً ضمن مجال الرؤية. وهذا هو مشروع واحد فقط في ترسانة المشاريع الاستخباراتية المماثلة وذات الصلة.

وبالمثل، فيما يتعلق بالمشاريع العلمية، يتكوّن تسلسل الجينوم البشري من نحو ١٠٠ غيغابايت من البيانات (Vanacek 2012): وهذا يتضاعف عبر الآلاف من الأفراد مما يوسع قاعدة البيانات قريباً إلى تيرابايت وبيتابايت من البيانات. وعندما بدأ التشغيل في مسح سلوان الجوي الرقمي (Sloan Digital Sky Survey) في العام ٢٠٠٠، أنتج

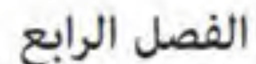


المرقاب (التلسكوب) في نيومكسيكو من بيانات الرصد والمراقبة في أول شهرين أكثر مما سبق جمعه في تاريخ علم الفلك حتى ذلك الوقت (Cukier 2010). وفي عام ٢٠١٠ م، بلغ حجم أرشيفها ما يقارب ١٤٠ تيرابايت من البيانات، وهي كمية سيتم قريباً جمعها كل خمسة أيام من قبل المرقاب الكبير للمسح الشامل (Large Synoptic Survey Telescope) والذي سيبدأ العمل به في شيلي في العام 2016 م(Cukier 2010). وأكثر ضخامة من هذا، مصادم هادرون الكبير (Large Hadron Collider) والذي يقوم بإنتاج ٤٠ تيرابايت من البيانات في الثانية الواحدة (The Economist, 2010). وفي هذا، وغيره من الحالات، يتم إنشاء بيانات هائلة للغاية لا يمكن تحليلها أو تخزينها، بحيث تعد بيانات عابرة. وفي الواقع، لا وجود لهذه القدرة التخزينية لكل هذه البيانات، فعلى الرغم من التوسع السريع في القدرات التخزينية، إلا أن هذا التوسع لا يمكنه مواكبة إنتاج البيانات (Gantz et al. 2007; Manyika et al. 2011).

الشمول (Exhaustivity):

تستخدم عملية انتقاء العينات في دراسات البيانات الصغيرة من أجل إنتاج مجموعة بيانية تمثيلية لمجتمع الدراسة من جميع البيانات المحتملة في وقت معين ومكان معين. وتستخدم عملية انتقاء العينات من مجتمع الدراسة بسبب كبر هذا المجتمع وتعذّر جمع كافة البيانات المتعلقة به لمحدودية الوقت والموارد المخصصة. في المقابل، تسعى مشاريع البيانات الكبيرة إلى التقاط بيانات كامل مجتمع الدراسة (ن = الكل)، أو على الأقل، حجم عينة أكبر من المعتادة في دراسات البيانات الصغيرة (Mayer-Schonberger and Cukier 2013). فمن ناحية، تُعد هذه البيانات نتاجاً ثانوياً للتقنيات التي استخدمت في إنتاج البيانات، جنباً إلى جنب مع النمو الهائل في القدرة على تخزين البيانات (انظر الفصل الخامس)، ومن ناحية أخرى، هناك قناعة بأن "الأكثر هو أفضل" وسيقدم تمثيلاً أكبر ومزيداً من الدقة والصحة في التحليل.

وفي النظم المفتوحة مثل المشاريع العلمية الكبيرة، كتلك التي تقيس البيانات المناخية للإبلاغ عن الطقس ونمذجة الأرصاد الجوية، أو جمع البيانات الفلكية باستخدام مرقاب قوي، يكون التوجه نحو مجموعات بيانات أكبر بكثير، مع زيادة أحجام العينات عبر العديد

مــن المتغيرات ما أمكن ذلك. ولا يعني ذلك في علم الفلك، على سـبيل المثال، مجرد جمع بيانـات الضوء، ولكن أيضـاً البيانات من مختلف ألوان الطيـف الكهرومغناطيسي، وبدقة عاليــة لأكبر قــدر ممكن. أما في حالة النظم المغلقة، مثل الفيسـبوك أو شراء السـلع من متجر على شـبكة الإنترنت مثل أمازون أو إرسـال رسـائل البريد الإلكـتروني، فمن الممكن تسـجيل كافة التفاعلات والعمليات التي تحدث، بالإضافة إلى مسـتوى عدم التفاعل. وفي هذه الحالات، هذا هو الواقع بطبيعة الحال، فموقع التواصل الاجتماعي الفيسـبوك يقوم بتسجيل كل نشر، أو تسجيل لإعجاب، أو رفع لصورة يتم تحميلها، أو وصلة إلى موقع آخر، أو رسـالة مباشرة، أو لعبة لُعبت، أو فترات الغياب، وما إلى ذلك لكل مسـتخدميه الذين ناهزوا المليار أو نحو ذلك. وبالمثل، لا تسجّل شركة الأمازون فقط تفاصيل كل عملية شراء والمشـترين، ولكن أيضاً جميع الروابط التي تمّ زيارتها وكل السـلع التي تمّ استعراضها من خلال موقعها، بالإضافة إلى كافة البنود التي تمّ وضعها في سـلة التسـوق ولم يتم شراؤها. كما يتم تسـجيل كافة رسـائل البريد الإلكتروني من قبل الخادمات التي تستضيف صندوق البريد الإلكتروني للعميل، وتخزين كامل البريد الإلكتروني وجميع البيانات الوصفية المرتبطة به على سـبيل المثال، الذين تم إرسـال البريد الإلكتروني لهم أو اسـتلامه منهم، والوقت / التاريخ، والموضوع، والملفات المرفقة. وحتى لو تمّ تحميل البريد الإلكتروني محلياً على جهاز المسـتخدم وتمّ حذفه فما زالت نسخة محفوظة منه على الخادم، هذا مع احتفاظ معظم المؤسسات والشركات بهذه البيانات لعدد من السنوات.

وبالمثل بالنسبة للأشكال الأخرى من البيانات، فقد نمت البيانات المكانية بشكل كبير في السـنوات الأخيرة، وذلك من خلال الاستشعار الآني عن بُعد والصور الرادارية، إلى المشاريع الكبـيرة مكتظة المصـادر مثل مشروع خرائط الشـوارع المفتوحـة (OpenStreetMap)، والمسـارات المكانية الرقمية التي تم إنشـاؤها بواسطة مسـتقبلات أنظمة تحديد المواقع الجغرافيـة المضمّنة داخل الأجهزة. ويهدف الصنفان الأوّلان إلى أن يكونا شـاملين مكانياً، بحيث تجمع التضاريس للكوكب الأرضي بأسره، ورسـم خرائط البنية التحتية لدول بأكملها وتوفير مجموعات بيانات مشتركة ومرخّصة للخرائط. في حين يوفّر المشروع الثالث القدرة على تتبّع وتعقّب الحركة عبر الفضاء بمرور الوقت؛ وبناء مسارات الزمن – الفضاء التي يمكن تجميعها لتوفير نماذج الزمن – الفضاء للسلوك عبر مدن ومناطق بأكملها. وتمكّن مجتمعة



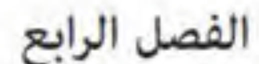
مــن النمذجة المفصّلة للأماكن وطرق التنقل، والمقارنة المســحية عبر الفضاء، واســتهداف التســويق لمجتمعات معينة، وتقديم خدمات جديدة معتمدة على الموقع، والبيانات التي يمكن الجمع بينها لاشــتراكها في المرجعيات المكانية لإنشاء قواعد بيانات وتطبيقات جديدة يمكــن البحــث مكانياً فيها (على ســبيل المثال، جمع البيانات حول منطقة لإنشــاء ملفات تعريف للحي).

وبالنظر إلى التقدم الحاصل في سعات التخزين (انظر الفصل الخامس)، يبدو أننا قد وصلنا إلى مرحلة بحيث يكون من الأسهل في كثير من الحالات تسجيل كل شيء، دون الحاجة إلى فرز البيانات، وتدقيقها، وأخذ عينات منها، وتســجيل ما يحتمــل أنه مفيد منها فقط (ومن الذي يعرف ما يثبت احتمال أنه مفيد في المستقبل؟). وكما أشار (2012) .Zikopoulos et al: «لا عجب أننا نغرق في البيانات، فإذا كنا نســتطيع تتبع وتســجيل شيء ما، فإننا بالعادة نفعل». وفي الواقع، أشار (2012: 7) Dumbill إلى أن المبدأ الأساسي خلف البيانات الكبيرة هو «عندما تســتطيع، احتفظ بكل شيء». ويتجلى هذا بالاعتقاد بأنه كلما كان المزيد من البيانات متاحاً، ســنحت فرص أفضــل لصنع رؤية أكثر صحة ونفاذاً... وفرص العثــور على «المنتجين» لنظرية جديدة (John Seely Brown مقتبس من 8 :2010 Bollier). وعلى الرغم من ذلك، تسهم إســتراتيجية السعي نحو الشمولية بشــكل كبير في طوفان البيانات، ويتمثل التحدي الأكبر في رؤية النمط الشــامل والصــورة الكبيرة من جملة التفاصيل وتمييز الغث من الســمين، وإثارة مجموعة من الأســئلة الأخلاقية فيما يتعلق بنطاق البيانــات التي يتم إنتاجها والاحتفاظ بها، والاستخدامات التي يجري توظيفها فيها أو من الممكن توظيفها فيها (انظر الفصل العاشر). كما أنها تثير تساؤلات معرفية أساسية (Floridi 2012). على سبيل المثال، وبالنظر إلى شموليتها، تساءل (2012) Callebaut عما إذا كانت تحليلات البيانات الكبيرة هي علم ما بعد الاختزال (Post-reductivist)، وسيتم مناقشة هذه الأسئلة بمزيد من التفصيل في الفصل الثامن.

الدقة والتبويب/ الفهرسة (Resolution and Indexicality):

بالإضافة إلى شــمولية البيانات، أصبحت البيانات الكبيرة أكثر تناهياً في الدقة بكثير، وهاتــان الخاصيتان يدفعــان جميعاً في الاتجاه نحو التبويب/ الفهرســة القوية (عنونة بوســم مميز وتحديد للهوية) (Dodge and Kitchin 2005)، وتعد صور الاستشــعار





عن بُعد مثالاً على تعزيز الدقة، ففي أواخر الثمانينيات، كانت أعلى دقة للصور المتاحة لمعظـم الباحثين غير الحكوميين عن سـطح الأرض هي لتلك الصور التي التقطتها أقمار لاندسات (Landsat Satellites) الصناعية، حيث تمثل كل نقطة ضوئية (بكسل) قطعة مـن الأرض تقـدر٣٠ x ٣٠ متراً، في حين أن الكثير من الصـور المتاحة على جوجل إيرث ذات دقـة تصـل إلى ٢,٥ x ٢,٥ مـتر، مما يتيح عـرض المزيد من التفاصيـل وتحليلها. وبالمثـل، فيما يتعلق بمخرجات بيانات التعداد السـكاني، فقـد زاد علو درجة البيانات في العديـد من الـدول. ففي الحالة الأيرلنديـة، كان نشر بيانات التعـداد حتى وقت قريـب للدوائـر الانتخابية (٣,٤٠٩ منطقة بمتوسـط ١,٣٥٠ نسـمة في كل منها، حيث كانـت الكثافة السـكانية لكل دائرة انتخابيـة أعلى بكثير في المـدن والبلدات وأقل في المناطق الريفية). وفي العام ٢٠١١، تمّ وضع بيانات التعداد بتوزيع جغرافي جديد سُمّي بالتجمعات الصغيرة، التي بلغ عددها ١٨,٤٨٨. تمثّل الوحدات الجديدة بالعادة بيانات من ٨٠ – ١٥٠ من الأسر (Gleeson et al. 2009). وقد أتاحت التجمعات الصغيرة إجراء تحليل التعداد على مسـتوى الحي أو الشـارع، بدلاً من مساحات شاسـعة جداً، ومناطق ذات أعداد متسـاوية تقريباً من السكان التي يمكن مقارنتها، مما يوفّر فهماً أكبر للتجمعات السـكانية والاقتصاد الأيرلندي. وإمعاناً في تحرّي المزيد مـن الدقة، يجمع العديد من وكلاء / سـماسرة البيانات كميات كبيرة من البيانات المتعلقة بالأفراد والأسر التي تمكن الشركات من استهدافهم بشكل فردي بالسلع والخدمات (انظر الفصل الثاني).

وقد صاحبت الزيادة في دقة البيانات، الزيادة في تحديد الهويات للأشخاص، والمنتجات، والمعامـلات، والأقاليم بحيث أصبحت بياناتها أكثر تبويباً وفهرسـة بطبيعتها (انظر الفصل الخامس). فعلى سبيل المثال، تحمل معظم السلع المعروضة للبيع في متاجر التجزئة الكبرى في الوقت الحاضر الرمز الشريطي (الباركود)، ويميّز هذا الرمز المنتج، ولكن ليس كل وحدة منه على حدة – حيث تشترك كل زجاجات صابون شعر الرأس (الشامبو) من نفس العلامة التجاريـة (الماركـة) في نفـس الرمز – مما يعني أنـه لا يمكن التمييز بينها بشـكل فردي. وفي المقابل، تتميز زجاجة صابون شـعر الرأس الموسومة بشريحة تحديد الهوية باستخدام موجـات الراديـو (Radio Frequency Identification Device- RFID) بشـكل فريد، لأن لكل شريحة رمز فريد من الممكن قراءته عن بُعد بواسـطة جهاز استشـعار الذبذبات



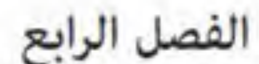
الراديوي. ونتيجة لذلك، يمكن تتبع كل زجاجة من مكان صنعها من خلال سلسلة التوريد وحتى المتجر وسلة العميل، مما يوجد مسار للتدقيق المفصل. وبعبارة أخرى، فقد أصبح من الممكن تتبّع الأمور الفردية بدقة عبر الزمان والمكان، بما في ذلك أولئك الذين يعملون على نقل أشياء من مكان لآخر. وبالمثل، بالنسبة للمعلومات، وخاصة تلك التي في شكل رقمي، حيث يتم تعريفها بشكل فريد من خلال رموز إدارة الحقوق الرقمية (Digital Rights Management). فعلى سبيل المثال، يمكن تعيين معرفات الكائن الرقمي (Digital Object Identifiers- DOIs) إلى الأعمال الإبداعية المتوفرة على الإنترنت (كالتقارير، المجلات، الصور، الصوت، وملفات الفيديو)، فمعرفات الكائن الرقمي هي محدد هوية دائم مرتبط بالبيانات الوصفية، مثل عنوان الإنترنت (Universal Resource Locator- URL) الذي يشير إلى موقع الملف. ويعزّز استخدام المعرّفات الفريدة العلائقية والقدرة على ربط وضم البيانات معا، كما يوفّر وسائل عملية لفرز ورصد ومطابقة وتمثيل الكائنات (Lyon 2003a; Dodge and Kitchin 2005; Graham 2005؛ انظر أيضاً الفصل العاشر).

العلائقية (Relationality):

تتمثل العلائقية بمدى القدرة على ضمّ مجموعات بيانات مختلفة بروابط بعضها مع بعض وكيف يمكن استخدام هذه الروابط للإجابة عن أسئلة جديدة. فالعلائقية هي في صميم قواعد البيانات العلائقية (انظر الفصلين الثاني والخامس)، وهي القدرة على إنشاء بيانات غاية في العلائقية التي تدفع بسوق البيانات الواسع وأرباح وكلاء البيانات وشركات الملفات التعريفية (انظر الفصل الثاني). إن الدرجة العالية من العلائقية هي ما يجعل التعداد السكاني مفيداً جداً لفهم التركيبة السكانية لأي أمة، وكيف تتغير عبر الزمان والمكان. وتختلف دراسات البيانات الصغيرة في مدى علائقيتها، أي ميل تلك التي تنطوي على البيانات المنظمة لدرجات أعلى من الترابط من تلك البيانات غير المنظمة، وهذا يستدعي القول إنه ينبغي وجود شكل من أشكال العلائقية بين البيانات لاستخلاص التفسيرات والاستنتاجات الشاملة منها.

وعلى الرغم من أن البيانات الكبيرة لا تستخدم غالبا بنية قاعدة بيانات علائقية (انظر الفصل الخامس)، فإن ميزة أساسية في طبيعتها هي العلائقية القوية. وكما فصّل





(boyd and Crawford (2011: 2 فإن «البيانات الكبيرة مترابطة بشكل أساسي، وتنبع قيمتها من الأنماط التي يمكن استخلاصها من خلال صنع روابط بين أجزاء من البيانات، عن فرد، أو أفراد على علاقة بآخرين، أو مجموعات من الناس، أو ببساطة عن هيكل تنظيم المعلومات نفسها». إلا أنها وخلافاً لقواعد البيانات العلائقية، فهي تتقن بالتساوي التعامل مع البيانات غير الرقمية (الرمزية)، والبيانات غير المنظمة كما البيانات المنظمة، والربط بين النوعين معاً وتعظيم القيمة المستفادة من التقاطعات بينها. ومن ثم، أصبح من الممكن الربط بين مجموعات متنوعة من البيانات – الشخصية، والعملية، والتفاعلية، والاجتماعية، والمالية، والمكانية، والزمنية، ...إلخ - وتحليلها على أساس فردي وجماعي للعلاقات والأنماط.

وقد ثبت جلياً، كمثال على قوة هذه العلائقية، الاستخدام المكثف للبيانات الكبيرة في الحملات الانتخابية للرئيس الأمريكي أوباما في العام ٢٠٠٨م و٢٠١٢م. فعلى النحو الذي فصّله (Issenberg (2012، سعى فريق أوباما لتحديد وتتبع جميع جوانب حملاتهم الانتخابية في ٢٠٠٨ م و٢٠١٢ م، ووضع سلسلة كاملة من المقاييس التي تم تسجيلها والتنقيب فيها بشكل مستمر للحصول على معلومات، وأنماط واتجاهات مفيدة. واشتمل ذلك على مراقبة صارمة للإجراءات الخاصة بهم، مثل وضع الإعلانات عبر وسائل الإعلام المختلفة، وإجراء حملات بريد الإلكترونية، والاتصال الهاتفي بالناخبين المحتملين، وطرق الأبواب وفرز الاصوات الانتخابية للمناطق، وتنظيم الاجتماعات والمسيرات، وتتبع الذين كانوا قد تحدثوا إليهم وماذا قالوا أو ما التزموا به. وقد عزّزوا هذه المعلومات بمئات التجارب العشوائية واسعة النطاق المُصمّمة لاختبار فعالية مختلف طرق إقناع الناس على التصويت لصالح أوباما أو التبرع بالأموال لحملته. وجمع فريق أوباما بين جميع المعلومات التي نتجت فيما تعلّق بالناخبين وبيانات التسجيل، والتعداد السكاني، والبيانات الحكومية الأخرى، ومسوحات الاقتراع، وبيانات جرى شراؤها من طيف واسع من الموردين، بما في ذلك وسطاء البيانات ووكالات التصنيف الائتماني، وشركات التلفاز الخطي (Cable TV). وكانت النتيجة مجموعة من قواعد البيانات الضخمة عن كل الناخبين في البلاد تتكون مما لا يقلّ عن ثمانين من المتغيرات (Crovitz 2012)، وأكثر من ذلك، أنها ترتبط بالخصائص الديموغرافية للناخبين المحتملين، وتاريخ انتخابهم، وكل حالة من حالات تواصلهم مع



حملة أوباما وردود أفعالهم، والتاريخ الاجتماعي والاقتصادي، وأنماط سلوكهم واستهلاكهم، ووجهـات النظر والآراء التي تمّ الإعـراب عنها، مع تحديث قواعـد البيانات يومياً خلال الحملـة بكل بيانات جديدة يجـري إنتاجها أو شراؤها. إما في الحالات التي لا يعرف فيها محللو أوباما الانتماء السياسي للناخب، ولا يمكنهم فيها الوصول إلى ذلك من خلال الاتصال المباشر معـه، فإنهم يوظّفون خوارزمية متطورة تسـتخدم المتغـيرات التي بحوزتهم عن شـخص ما للتنبؤ عن التصويت المحتمل المفضّل له (Issenberg 2012). وكانت النتيجة مليارات من أجزاء البيانات المترابطة والتي اسـتخدمت للتعريف بالناخبين بشـكل فردي، وتقييـم مدى احتمالية قيامهم بالتصويت والكيفية التـي يصوتون بها، والكيفية التي قد يسـتجيبون بها لمختلف السياسـات والقصص. لقد أنشـأ الترابط بـين البيانات في حملة أوباما مـا اصطلح عليه (Crampton et al. 2012) بمصطلـح «تأثير تضخيم المعلومات» (Information Aplifier Effect)، حيث مجموع البيانات أكثر من الأجزاء.

السرعة (Velocity):

هناك فرق جوهري بين البيانات الصغيرة والبيانات الكبيرة يتمثل في الطبيعة الديناميكية لإنتاج البيانـات، فالبيانات الصغيرة تتكوّن عادة من الدراسـات التي تكون مؤطرة بقالب جامد من حيث الزمان والمكان. وحتى في الدراسات المطوّلة، يتم التقاط البيانات في أوقات متفرقة ذات فاصل زمني (على سبيل المثال، كل بضعة أشهر أو سنوات). فعلى سبيل المثال، يتم إجراء التعدادات السـكانية بشـكل عام كل خمس أو عشر سـنوات، وفي المقابل يتم إنشـاء البيانات الكبيرة على أسس أكثر استمرارية بكثير، وفي كثير من الحالات بشكل آني أو لحظي وبشـكل مباشر بالتقارب من الوقت الحقيقي للوقوع (Real-Time) . وعوضاً عن قدر ضئيل ومتفرق من البيانات، التي تُجمع أو تعالج بمشـقة، فإن البيانات الكبيرة تتدفق بسرعـة، ولذلك كان هناك انتقال في التعامل من المعالجة بالدفعات (Batch processing) إلى البيانـات المتدفقـة (Streaming Data) (Zikopoulos et al. 2012). فمـن ناحية، يسـهم هذا في مسـألة كبر حجم البيانات من خلال إنتاج بيانات بسرعة أكبر، ومن ناحية أخرى، فهو يجعل كامل دورة البيانات أكثر ديناميكية بكثير، ويبرز قضايا كيفية إدارة نظام البيانات المتغير باستمرار.

وتنبع السرعة من التكرار المستمر لتسجيل المشاهدات مع مرور الوقت و/أو تغيّر المكان (Jacobs 2009) مع كون العديد من الأنظمة في وضع التشغيل الدائم باستمرار (Dodge and Kitchin 2005). فعلى سبيل المثال، تسجل المواقع الإلكترونية بيانات كافة الزيارات التي تمت للموقع والأنشطة التي قام بها الزائر خلال تصفحه على الموقع؛ كما تقوم المعدات الطبية بمراقبة العلامات الحيوية باستمرار، وتسجّل الكيفية التي يستجيب بها الجسم للعلاج وتُطلق إنذاراً في حال تجاوزت هذه العلامات حدّ معين؛ وتتعقب شركات الهاتف المحمول موقع، وهوية، وطبيعة استخدام الأجهزة الموصولة على شبكاتها كل بضع ثوان؛ وبالمثل ترصد شبكات استشعار الطقس مؤشرات الغلاف الجوي كل بضع دقائق وترسل نتائجها إلى قاعدة بيانات مركزية لإدراجها في الاستطلاعات الجوية؛ ويسجّل جهاز استشعار الذبذبات الراديوي (Transponder) على طول طرق المدينة البرية والسكك الحديدية هوية الحافلات والقطارات التي تمرّ، مما يتيح لسلطة النقل العام معرفة أماكن مركباتها في أي وقت، واحتساب الوقت المقدّر للوصول في محطات التوقف المختلفة؛ كما تراقب متاجر التجزئة مبيعات آلاف المنتجات المختلفة للآلاف من العملاء، حيث تستخدم البيانات لمعرفة أوقات إعادة تعبئة الرفوف وطلب دفعات جديدة من الموردين؛ ويتواصل الناس بعضهم مع بعض من خلال مواقع وسائل التواصل الاجتماعي في سيل لا ينتهي من التبادلات والعلاقات البينية المتشابكة؛ ويرصد المرقاب (Telescope) باستمرار السماء لقياس التقلبات في موجات الراديو من أجل فهم أفضل لطبيعة الكون، ففي كل هذه الحالات، هناك تيار مستمر من البيانات التي تتطلب إدارة وتحليلاً مستمراً.

ويشكّل نقل وإدارة كميات كبيرة من البيانات المنتجة باستمرار تحدياً تقنياً بسبب التحديات الناشئة عن محدودية قدرة سعة التخزين مما يوجد اختناقات بشكل سريع في حركة مرور وتبادل البيانات. على سبيل المثال، وكما أن فيديوهات موقع اليوتيوب (YouTube) قد تتوقف أو يتأخر عرضها بسبب كون سعة قناة الاتصال بالإنترنت ليست كافية لتواكب سرعة تدفق البيانات المطلوب، فإن الأثر نفسه يمكن أن يقع في حالة جمع ومعالجة البيانات، حيث تكون الأنظمة عاجزة عن مواكبة تدفق البيانات. وتشمل حلول هذه المشكلة زيادة سعة قناة الاتصال، وتوظيف تقنيات فرز وضغط البيانات مما يقلل من حجمها ومن ثم وقت معالجتها، وتحسين كفاءة خوارزميات المعالجة وتقنيات إدارة

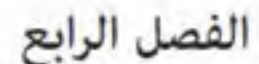


البيانات. كما يمثّل تحليل سـيل البيانات المتدفقة تحدياً أيضاً، لأن النظام لا يسـكن في أي لحظة وعند أي نقطة، كما في حالات الأسواق المالية، حيث يمكن أن يكون تحليل الصفقات لمايكرو ثانية من الزمن قيّماً للغاية. وهنا، تستخدم خوارزميات معقّدة، إلى جانب العرض المرئي للبيانات الديناميكية وهي في حالة تغير مستمر، لتتبع وتقييم النظام.

التنوع (Variety):

قد تختلف البيانات الصغيرة والبيانات الكبيرة في طبيعتها، كونها منظمة، أو غير منظمة، أو شـبه منظمة، وتتألف من الأرقام، النصوص، الصـور، الفيديو، وغيرها من أنواع البيانات الأخـرى. إلا أن هذه الأنـواع المختلفة من البيانات يرجح أن تكون في البيانات الكبيرة أكثر قابليـة للجمع والربط بينها، وإمكانية وصل البيانات المنظمة بالبيانات غير المنظمة. فعلى سـبيل المثال، تتألف نشرات الفيسبوك من النصوص التي ترتبط في الغالب بصور أو ملفات فيديـو أو غيرها من المواقع، وهي تجتذب تعليقات مسـتخدمي الفيسـبوك الآخرين؛ كما قـد تجمع شركة ما بين البيانات المالية المتعلقة بالمبيعات واسـتطلاعات العملاء التي تعبّر عـن مدى ثقتهـم بالمنتج، وفي المقابل تكـون البيانات الصغيرة أكثر انفصـالاً وترتبط، على كل حـال، من خلال مفاتيح تعريفية وحقول بيانات مشـتركة. ويتمثّل السـبق الأسـاسي بالنسبة للبيانات الكبيرة في كيفية اختلافها عن الأشكال التي سبقتها لإدارة البيانات الرقمية، والتي كانت احترافية للغاية في معالجة وتخزين البيانات الرقمية باستخدام قواعد البيانات العلائقية، مما مكّن من إجراء مختلف أنواع التحليل الإحصائي. إلا أنها كانت أضعف بكثير في التعامـل مع نماذج البيانـات الغير رقمية، عدا عن مجرد تخزينهـا في ملفات مجردّة أو مضغوطة. وكما لاحظ اتحاد مراكز البيانات المفتوحة Data Center Alliance (2012: 7) بأنه في السـابق كان يجري تجاهل البيانات غير المنظمة، أو في أحسـن الأحوال، استخدامها بشـكل يفتقر إلى الكفاءة. ومع ذلك، فإن التقدم الحاصل في الحوسـبة الموزعة، وتصميم قواعـد البيانات باسـتخدام هياكل عوضاً عن لغات الاسـتعلام الهيكلية (NoSQL) (انظر الفصل الخامس)، والتنقيب في البيانات وتقنيات اكتشـاف المعرفة (انظر الفصل السادس)، قد زادت بشـكل كبير القـدرة على إدارة ومعالجة واسـتخراج المعلومات من البيانات غير المنظمة. وفي الواقع، يجري على نطاق واسع تداول أن ما يقارب ثمانين في المائة من جميع





البيانــات الكبيرة هي غير منظمة بطبيعتها، وكــما أبان (Grimes (2011)، فقد أصبح هذا الرقم مسلّماً به مع قليل من الإثبات.

المرونة (Flexibility):

يكــون تصميم البحث وإدارة البيانات، مع مشــاريع البيانات الصغيرة، غير مرن نســبياً بمجــرد بدء العمل الميــداني والتحليل، وذلك بالنظر إلى الخدمات اللوجســتية، والتكاليف، والحاجة إلى التمثيل في عينة صغيرة الحجم. فعلى سبيل المثال، فإنه من الضروري على كل شــخص تمّ جمع بيانات التعداد له تعبئة النموذج ذاتــه للتأكد من إمكانية قابلية مقارنة البيانات عبر كافة الســكان. وبمجرد طباعة النماذج، لا يمكن إضافة حقول بيانات جديدة، مما يعني ثبات البيانات الممكن استخراجها من هذه النماذج، وبالمثل تميل قواعد البيانات العلائقيــة التــي تُحفظ بها البيانات إلى أن تكون قالباً ثابتــاً ومحدودة النطاق. وعلى نحو مماثل، يكون تصميم البحث عادة، في التجارب العلمية والدراسات البيئية، غير مرن بمجرد بدئهــا، وذلك بهدف تمكين إجراء عمليات المقارنة والتكرار. أما في الدراســات التي تعتمد المقابلات أو الأنشطة الاجتماعية، فمن الممكن للباحث، على الرغم من ذلك، أن يكون أكثر مرونة في النهج المتّبع، وأن يكون لديه نموذج حرّ من الأسئلة وإمكانية التكيف مع الأوضاع المستجدة. ويمكن أن يكون ترميز وإدارة وتحليل هذه البيانات مرناً نسبياً، ولكن هذا يعود جزئياً إلى محدودية حجم ونطاق مجموعة البيانات.

في المقابــل، تمّ تصميم نظم البيانــات الكبيرة لتكون مرنــة بطبيعتها، بحيث تحمل ســمات قابلية الاســتطالة والتفرع (يمكن إضافة حقول بيانات جديدة بسهولة) وقابلية التوســع (يمكن التوســع بسرعة) بغض النظر عن الحجم (Marz and Warren 2012). ويعني استخدام قواعد البيانات غير الهيكلية للاستعلام (NoSQL) إمكانية إدارة البيانات المتغــيرة بسرعة عاليــة، والتكيف مع حقول البيانات الجديــدة (انظر الفصل الخامس)، وهذا يعني أنه من الممكن التكيّف مع إنتاج البيانات على أسس تدويرية، وأداء الاختبار عــلى التكيف. فعلى ســبيل المثال، تقــوم شركات جوجل، والفيســبوك، ومنصات أخرى على الإنترنت بتعديل تصاميمها، وجمع بيانات حول كيفية اســتجابة المســتخدمين لهذه التغييرات على سبيل المثال، مراقبة الزيارات العابرة للموقع، وتحليل النتائج واستخدامها



لتقديم مزيد من التعديلات المحسنة للتصميم بغية تشجيع إجراءات وردود معينة بشكل مستمر. ولأن أعداد الأشخاص الذين يستخدمون هذه المواقع هائلة، وأحجام العينات الخاصة بهم كبيرة جداً، فهذا يعني أن بإمكانهم إجراء التغييرات دون الخوف من فقدان التمثيل. على سبيل المثال، وبالعودة إلى حملة باراك أوباما الانتخابية، فقد أجرى فريق حملته تجارب متتالية عن مدى فعالية التعديلات المحسنة لموقع الحملة (BarackObama.com) لزيادة المشاركة، والعمل التطوعي، والتبرعات. فأحد الاختبارات قيّم آثار التغيير على زر «اشترك» إلى «مزيد من المعلومات»، و «انضم إلينا الآن»، و«اشترك الآن»: وأصبح من الواضح على مدى ثلاثمائة ألف زيارة أن استخدام زر «انضم إلينا الآن» قد أدى إلى عشرين في المائة زيادة في الزائرين المقبلين على التسجيل في الموقع (Issenberg 2012).

فمثل هذه المواقع الكبيرة الحجم ينبغي أن تكون أيضاً قابلة للتوسع والتطوير، وقادرة على التعامل مع الطفرات الناشئة في طلب وإنتاج البيانات، حيث يتسبب مقدار حركة مرور البيانات في انهيار قاعدة البيانات العلائقية التقليدية العاملة على خادم واحد. فعلى سبيل المثال، يمكن أن يزداد عدد التغريدات التي يتعامل معها تطبيق التواصل الاجتماعي تويتر بشكل ملحوظ، مع عشرات الآلاف من التغريدات التي يجري نشرها كل بضع ثوان أثناء الأحداث الكبيرة، مثل حفل افتتاح الألعاب الأولمبية أو خلال المباراة النهائية. وكان الحل هو تركيب أنظمة عتاد تتألف من أجزاء موزّعة بحيث تخزّن البيانات في قواعد بيانات مقسّمة عبر العديد من الخوادم، مما مكّن من توسيع نطاق التخزين بحسب الحاجة. وعلاوة على ذلك، ففي بعض الأنظمة، مثل تويتر، يمكن تعيين المرونة من قبل المستخدمين لتحديد ما إذا كانت تشمل البيانات أم لا. على سبيل المثال، يقرر المستخدمون في العديد من تطبيقات وسائل التواصل الاجتماعي والمتنقلة إدراج أو عدم إدراج مواقعهم، وكذلك البيانات التعريفية الرئيسية الأخرى المتعلقة بالهوية (Gorman 2013).

الخلاصة:

البيانات الكبيرة هي ظاهرة حديثة، وبالنظر إلى التطبيق والانتشار المتسارعين لها، فقد برزت نقاشات مستمرة حول ما يشكّل البيانات الكبيرة والخصائص المرتبطة بها. وتُعد بعض التعاريف، مثل تلك التي تنص على أن البيانات الكبيرة هي أي مجموعة بيانات كبيرة جداً بالقدر الذي لا يمكن معه تخزينها في جدول بيانات إكسل أو على جهاز واحد (Strom 2012)، مبتذلة للغاية وغير مفيدة، حيث تخفض مستوى مفهوم البيانات الكبيرة إلى خاصية الحجم فقط. لقد أصبح من الواضح أن لدى البيانات الكبيرة عدد من الخصائص الملازمة لها والتي تجعلها مختلفة نوعيا عن الأشكال السابقة من البيانات، وتمّ في هذا الفصل مناقشة الخصائص الأساسية السبعة للبيانات الكبيرة: الحجم، والسرعة، والتنوع، والشمول، والدقة / الفهرسة، والعلائقية، والمرونة / قابلية التوسع التي تميزها عن البيانات الصغيرة (انظر الجدول ٢-١).

فهذه هي دورة المستوى الأول فيما يتعلق بتقديم تقييم توصيفي لطبيعة البيانات الكبيرة. وهناك حاجة إلى مزيد من العمل لتقييم البيانات الكبيرة الناتجة عن مصادر متعددة لتحديد ما إذا كان هناك أصناف متنوعة في طبيعة البيانات الكبيرة. فعلى سبيل المثال، قد تحمل بعض البيانات خمساً أو ستاً من هذه الخصائص، ولكنها لا تستوفي أو ضعيفة في واحدة أو اثنتين، وقد تفتقر مجموعة بيانات إلى التنوع (منظمة بشكل كبير)، أو الحجم (صغيرة، ولكن شاملة مع حجم العينة ن = الجميع)، أو قليلة السرعة (يتمّ إنشاء البيانات بانتظام ولكن في كل شهر عوضاً أن تكون بشكل مستمر)، أو تفتقر إلى الفهرسة (كونها مجهولة المصدر أو مجمعة)، مع حملها للخصائص الأخرى على سبيل المثال. فمن الواضح أن مثل هذه البيانات ليست بيانات صغيرة كما هو مبيّن في الفصل الثاني، ولكنها في ذات الوقت ليست بيانات كبيرة كما تُفهم بمعناها الضيق الذي يصفها بكل الخصائص السبع، ولكن يمكن اعتبارها على الرغم من ذلك شكلاً من أشكال البيانات الكبيرة. وبعبارة أخرى، هناك حاجة لإنتاج تصنيف للبيانات الكبيرة قائم على أدلة تجريبية قوية مع أمثلة الحالات التي من شأنها أن تساعدنا على التفكير بصورة أكبر في طبيعة هذه البيانات. ويحتاج هذا إلى أن يكون مصحوباً بدراسة للخصائص الأخرى، مثل نوعية البيانات وصحتها، ودقتها، ومنشئها (انظر الفصل التاسع).



كما تثير الخصائص السبع للبيانات الكبيرة تساؤلات بشأن الآثار المترتبة على طوفان مثل هذه البيانات، وما الذي يعنيه القدرة على الوصول إلى بيانات كبيرة جداً، وشاملة، ومتغيرة، ودقيقة، ومفهرسة، ومتنوعة، وعلائقية، ومرنة، وقابلة للتوسع بالنسبة للمجتمع والحكومة ورجال الأعمال؟ وإلى أي مدى يمكن أن توفر هذه البيانات رؤى ثاقبة عن حالة الإنسان أو المساعدة على تناول بعض القضايا الاجتماعية، والسياسية، والاقتصادية، والبيئية الأكثر إلحاحاً والتي تواجه كوكب الأرض؟ أو، بدلاً من خدمة الصالح العام، فهل ستستخدم هذه البيانات لمزيد من خدمة المصالح الخاصة في الغالب؟ أو خدمة مصالح الدولة؟ وكيف سوف تغير هذه البيانات في نظرية المعرفة العلمية في جميع المجالات (الآداب والعلوم الإنسانية، والعلوم الاجتماعية، والعلوم الفيزيائية والحياتية، والهندسة)؟ فالفصل السابع والثامن والعاشر ستناقش هذه القضايا بالتفصيل، وتقدم التفكير النقدي لآثار وعواقب البيانات الكبيرة.







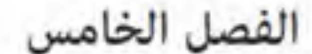
الفصل الخامس

مُمكنات البيانات الكبيرة ومصادرها

برز النمو السريع للبيانات الكبيرة نتيجة التطور المتزامن معه لعدد من التقنيات المساعدة، والبنى التحتية، والأساليب والممارسات، وتضمينها بشكل سريع في الأعمال اليومية، والممارسات، والمســاحات الاجتماعية. وتضمنت هذه البنية التحتية المعرفية الجديدة إطلاق طيف واســع مــن مجموعة متنوعة مــن تقنيات المعلومــات والاتصالات، وخاصة خدمــة الإنترنت الثابتة والمتنقلة عبر الهاتف المحمول؛ وإدخال البرمجيات في جميع الكائنات، والآلات والأنظمة، ناقلة إيّاها من «الغباء» إلى «الذكاء»، فضلاً عن إنشاء أجهزة وأنظمة رقمية بحتة بالكامل؛ وتطوير الحوســبة الشمولية الواسعة الانتشار والقدرة على الوصول إلى الشبكات والحوسبة في العديد مــن البيئات، وعلى ذات خطى هــذا التحرك؛ التقدم في تصميم قواعــد البيانات ونظم إدارة المعلومــات؛ وعروض التخزين الموزّع والتخزين الأبدي للبيانات بأســعار معقولة؛ والأشــكال الجديدة من تحليلات البيانات المصممة للتعامل مع وفرة البيانات بدلاً من ندرتها.

وإضافة إلى تمكين الوصول والمشــاركة والتخزين للبيانات، تعد البنية التحتية والمعرفية الجديدة في الغالب الوســيلة التي يتمّ إنتاج الكثير من البيانات الكبيرة منها. ففي الواقع، وبلا شــك أن الحالة في وقتنا الحاضر تتمثل في أننا نعيش في مجتمعات تســتخدم على نحو متزايد تقنيات البيانات المكثّفة والمنتجة لها وتنتظم على أساســها. كما أصبحت الممارسات الشــخصية في الاتصال، والاستهلاك، والإنتاج، والســفر، والحياة المنزلية، أكثر فأكثر تتوسّطها الأجهــزة الرقميــة التي تنتج الكثير من البيانات المســتنفذة القيّمة. وعلاوة على ذلك، يتم الآن مراقبــة وتنظيم الأماكن التي نعيش فيها على نحو متزايد بمجاميع كثيفة من التقنيات والبنى التحتية المدعّمة بالبيانات، مثل أنظمة المرور وإدارة المباني، ونظم المراقبة والشرطة، وقواعــد البيانات الحكومية، وإدارة العملاء وسلاســل الدعم اللوجســتية، والأنظمة المالية ونظم المدفوعات. وستُناقش العوامل المساعدة ومصادر البيانات الكبيرة بالتفصيل في هذا الفصل.





العوامل المساعدة للبيانات الكبيرة:

تُعـد البيانات الكبيرة مُحصلة للتطور والتقارب بـين مجموعة متقدمة من التقنيات في مجال الحوسبة منذ نهاية الحرب العالمية الثانية. ويشمل ذلك إنتاج أجهزة الحاسب المركزية في الخمسينيات والستينيات؛ وشبكات الإنترنت الوليدة في السبعينيات والثمانينيات التي ربطت الحاسبات بعضها ببعض؛ والانتشار الواسع النطاق للحاسبات الشخصية في الثمانينيات والتسعينيات؛ والنمو الهائل للإنترنت في التسعينيات، وتطوير الصناعات القائمة على الويب، جنباً إلى جنب مع النمو الهائل في الهواتف النقالة والأجهزة الرقمية مثل أجهزة الألعاب والكاميرات الرقمية؛ وتطوير الحوسبة السحابية، والموزعة، والنقّالة، وتقنيات الإصدار الثاني لشبكة الويب (Web 2.0) في بدايات الألفية الثانية؛ وتفشي الحوسبة الشمولية واسعة الانتشار (Ubiquitous and Pervasive Computing) خلال هذا العقد. وخلال هذه الفترة وقع عدد من المؤثرات التحويلية، فزادت القدرة الحاسوبية أضعافاً مضاعفة، وتمّ ربط الأجهزة معاً، وتداخلت الأنظمة الرقمية بازدياد متعاظم في شئون الحياة اليومية وعملياتها، وأصبحت البيانات أكثر من أي وقت مضى مفهرسة وقابلة للقراءة الآلية، كما توسعت أنظمة تخزين البيانات وأصبحت موزعة.

الحوسبة (Computation):

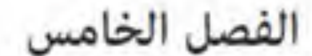
بينما قدّمت الحواسيب الرقمية المركزية الأولية في الخمسينيات والستينيات قدرة حاسوبية أكثر كفاءة من تلك التي عند البشر والأجهزة التناظرية التي استخدموها (مثل المعداد، والآلات الحاسبة الميكانيكية، والآلات الحاسبة بالبطاقات المثقوبة، وأجهزة الحاسب التناظرية، إلخ)، إلا أن قوة المعالجة لها كانت محدودة، ومن ثم كانت أنواع العمليات التي يمكن أداؤها مقيدة، وعلاوة على ذلك كانت تلك الحواسيب كبيرة الحجم ومكلفة. على سبيل المثال، كان بإمكان حاسب الإينياك (Electronic Numerical Integrator Analyser and Computer- ENIAC)، الذي طُوّر في منتصف الأربعينيات، أداء خمسة آلاف عملية جمع وطرح في الثانية، ولكن اقتصرت ذاكرته الرئيسية عالية السرعة على عشرين كلمة (أي ثمانين بايت)، لآلة تزن ثلاثين طناً، وتحتوي ١٨,٠٠٠ أنبوب فراغي، و١٥,٠٠٠ مرحّل، ومئات الآلاف من المقاومات والمكثفات والمحاثات،



وباســتخدام مائتي كيلوواط من الطاقة الكهربائيــة (Najmi 2004). وفي العام ١٩٦٥م، توقّع Gordon Moore من إنتل بأن يتضاعف عدد الترانزســتورات في الدوائر المتكاملة كل ثمانية عشر شــهراً في العقد المقبل ذلك الحين، مما يزيد بشكل كبير القدرة الحاسوبية (وهو ما أصبح يُعرف لاحقاً باسم قانون مور). وقد أثبت هذا التوقّع بصيرة نافذة حيث تــمّ إدخال تحســينات تقنية على مكونات وتصميم أجهزة الحاســب (من خلال الانتقال مــن الأنابيب الفراغية إلى الترانزســتورات إلى الدوائر المتكاملــة إلى المعالجات المصغّرة)، وازدادت القوة الحســابية فيما يقرب من الخمســين في المائة ســنوياً بين عامي ١٩٤٠ م و٢٠٠١م (Nordhaus 2002). وإلى جانــب هذا النمو، انخفضت تكلفتها بشــكل كبير، حيث تشــير إحدى التقديرات بانخفاض في تكلفة الحســاب ناهز أضعاف الكوادريليون (١٠٠٠٠٠٠٠٠٠٠٠٠٠٠٠) خلال القرن العشرين (CRA 2003).

اســتمرت مثل هذه الزيادات الهائلة في القدرة الحاســوبية والانخفاضات الكبيرة في التكلفة في الألفية الجديدة، ولكن الأهم من ذلك أنها أصبحت متنقلة على نطاق واسع مع انتشار مكثف للأجهزة الحاسوبية المحمولة مثل أجهزة الحاسب المحمول، والأجهزة اللوحية والهواتف الذكية، وأجهزة الملاحة المرتبطة بالأقمار الصناعية (SatNav)، وغيرها. وتعادل قوة المعالجة للهاتف الذكي المتوسط "عشرة أضعاف قوة جهاز الحاسب المركزي كراي1- (Cray-1) الذي تمّ تشغيله في مختبرات لوس ألاموس الوطنية في عام 1976م " (Townsend 2013: xiii)، مع اســتخدامها جزء من قوة معالجتها، حيث تمتلك قدرات حاســوبية، وتخزينية، وتطبيقات أكثر من العديد من أجهزة الحاســب الشــخصية منذ عقد واحد فقط من الزمن، وهي أيضاً تشــغل جزء بســيط في حجم العتاد المســتخدم. على ســبيل المثال، لــدى العديد من الهواتف الذكية معالجا مصغــرا بسرعة ١,٥ - ٢,٠ غيغاهرتــز و١٦ - ٣٢ غيغابايت من الذاكرة الرئيســية. وتمكّن هذه القوة في المعالجة وســعة الذاكرة من معالجة كميات كبيرة جداً من البيانات بسرعة، ليتمّ تخزينها محلياً، لأنها متصلة شبكياً ومشتركة.





الربط (Networking):

بذات القدر من الأهمية للنمو الكبير في القوة الحسابية لمعالجة البيانات الرقمية كان الربط بين الحاسبات بعضها مع بعض والذي سهل نقل وتبادل البيانات فيما بينها. وقد وقع أول اتصال شبكي بين جهازي حاسب عبر شبكة الإنترنت في نوفمبر عام ١٩٦٩ م، وقد تم الربط بين نقطة في جامعة كاليفورنيا في لوس أنجيلوس (UCLA) ونقطة أخرى في جامعة ستانفورد (Salus 1995). ثم نمت الشبكة ببطء، من ١٣ نقطة في يناير ١٩٧١م، إلى ٢٣ في أبريل ١٩٧٢م، إلى ٦٢ في يونيو ١٩٧٤م، و١١١ بحلول مارس ١٩٧٧م، واقتصرت حينها على متعاقدي وكالة مشاريع البحوث المتطورة الدفاعية (Hart et al. 1992; O'Neill ARPA) 1995). وتعمل شبكة الإنترنت من خلال تقطيع البيانات والتعليمات في حزم متساوية الحجم ونشرها عبر النظام، مع عنونة كل حزمة بمعرّف وعنوان المرسل إليه المطلوب، ثم تمرّر كل حزمة من نقطة اتصال إلى أخرى إلى أن تصل إلى الوجهة المقصودة حيث يتم إعادة تجميعها هناك. ولتمكين اتصال الأجهزة والشبكات التي تعمل على سرعات مختلفة على مدار الساعة وتستخدم حزماً مختلفة الحجم، اعتمدت مجموعة معيار (بروتوكول) الشبكة تي سي بي / آي بي (TCP/IP) (معيار التحكم بالإرسال / بروتوكول الإنترنت) في عام ١٩٧٣ (O'Neill 1995)، ولا يزال هذا المعيار يستخدم حتى يومنا هذا. كما يؤكد هذا المعيار على حيادية الشبكة، حيث يتم التعامل مع كافة البيانات على قدم المساواة ودون تمييز بينها أو تفاضل في التكلفة (Wu 2011).

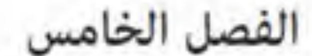
وبحلول أواخر السبعينيات، تمّ إنشاء الشبكات المحلية الداخلية، كما تمّ إنشاء الشبكات الأخرى دولياً وتمّ ربطها بشبكة الإنترنت الوليدة، إضافة إلى شبكات الهواة المستقلين والشركات، كما أصبح من الممكن إرسال حزم البيانات عبر الأقمار الصناعية. وبالإضافة إلى نقل البيانات والملفات، تم إنشاء البريد الإلكتروني واللوحات الإعلانية (Kitchin 1998). وخلال الثمانينيات، نمت البنية التحتية، مع لاعبين جدد من المؤسسات والشركات، جنباً إلى جنب مع تطور شبكات الإنترانت الداخلية (الشبكات الخاصة). وفي عام ١٩٩٢م اخترع تيم بيرنرز لي (Tim Berners-Lee) من سيرن (CERN) في جنيف الشبكة العالمية (الويب)، وقدّم وسيلة سهلة الاستعمال أكثر بكثير للوصول إلى الإنترنت واستخدامها. وخلال



التسعينيات وبدايات الألفية الجديدة، تمّ تطوير تقنيات شبكية جديدة مثل شبكات اتصال الميــدان القريب (NFC) والتواصل المباشر (Proximate Communication) باســتخدام معايــير البلوتوث، والتغطية المحلية اللاســلكية الــواي فاي (WiFi)، والشــبكات الوطنية للاتصالات المتنقلــة (GSM/3G). ووفقاً لقوانين جورج غيلدر (2000) (George Gilde) في الاتصــالات، فــإن إمدادات العالم من عرض النطاق الــترددي (قدرتها على نقل البيانات) تتضاعف كل ســتة أشــهر، مع مزيد من القدرة الإضافية التي تقدّمها الشبكات اللاسلكية. ونتيجة لهذه التطورات، فقد أصبح الربط متاحاً بين الأجهزة الحاســوبية من خلال شــبكة الإنترنت بشكل أسهل على نحو متزايد، وأسرع وعلى نطاق أوسع.

الحوسبة الشمولية واسعة الانتشار:

صاحب التوســع في النفاذية وقدرة الوصول إلى شــبكات تقنية المعلومات والاتصالات وزيادة ســعات قنوات الاتصال (عرض النطاق الترددي)، تنوّع وتوسّعٌ في الأجهزة الرقمية والتــي تتصل إما بشــكل مباشر أو غــير مباشر (حيث يتمّ تحميلها عــلى جهاز آخر أولا) بشــبكة الإنترنت لتبادل البيانــات، أو التعليمات، أو الحصول عــلى التحديثات البرمجية. ففي بعض الحالات، حلّت التقنيات الرقمية الجديدة محلّ الأشكال السابقة من البيانات التماثلية / التناظرية، فعلى ســبيل المثال، تمّ اســتبدال أنظمة الهاتف التماثلي بالهواتف المحمولة والهواتف الذكية، واستبدلت كاميرات المراقبة التناظرية التي تسجّل على شريط فيديــو بكاميرات رقمية تلتقط الصور الرقمية، ويجري الاســتعاضة عن الموســيقا المخزنة عــلى الشريط المغناطيــسي أو الفينيل بملفات صوتيــة رقمية، والنــماذج الورقية بأخرى إلكترونيــة على الإنترنت. من هنــا، تم تحويل الأجهزة والأنظمة التي كانت تُعد نســبيا «غبيــة» لتصبح «ذكية» ببعض الطــرق، وعادة من خلال اســتخدام البرامج المرتبطة بها (Dodge and Kitchin 2005). وتعنــي كلمة «الذكية» هنا بشــكل عام بأن الجهاز قد تمــت برمجته ليكون واعياً للطريقة الخاصة التي يســتخدم بها وقادراً على اتخاذ قرارات مســتقلة، وآنية، ومؤتمتة من مجموعة الخيارات المحددة مــن خلال تطبيق خوارزميات عــلى البيانات المنتجــة (انظــر Dodge and Kitchin 2007a). وخير مثال على تجميع تقني «غبي» أصبح على نحو متزايد «ذكي» هو الســيارة، فالســيارة الحديثة يتمّ تحميلها

بأجهزة رقمية، ومستشـعرات ومحركات، والتي تراقب وتنظم أنظمتها المختلفة وتتداخل مـع تجربة القيـادة (Thrift 2004). ويمكن أن تنتقل البيانات المخزنة من هذه الأجهزة عبر شـبكات الاتصـالات عن بُعد للـشركات أو تحميلها من قبل مهنـدس لإجراء تحليل تشـخيصي (Dodge and Kitchin 2007a). وفي حـالات أخرى، هناك أجهزة لا مكافئ تناظرياً لها، حيث تنتج بيانات جديدة كلياً في شكلها، مثل تلك الناتجة عن أنواع جديدة من التقنيات الرقمية، والمسـح الضوئي، وأجهزة الاستشـعار على سـبيل المثال، ماسحات التصويـر بالرنـين المغناطيسي MRI في مجـال الصحة، وإضافة إلى جعلهـا «ذكية»، فإن الأنظمة تعمل في الغالب على جمع أكبر قدر من البيانات قدر الإمكان، وتسـعي لتكون هذه البيانات شاملة وليست انتقائية. وهنا، قد يتمّ جمع بيانات عرضية أو غير ضرورية لتوفير الخدمة أو العملية، مثل التفاصيل الشـخصية أو الموقع الجغرافي. فغالباً ما تُحدد الإعدادات الافتراضية للأجهزة بشـكل يتيح للناس إمكانية اختيار الانسـحاب من أنظمة إنتاج البيانات من خلال عقود الاستخدام للبرمجيات والمواقع، مع النص على العقوبات إذا اختاروا القيام بذلك، مثل تردي الخدمات، والتكاليف الإضافية، أو حجب الخدمة بالكلية.

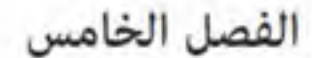
وبالمثـل، حصل هذا النمو في الأجهزة الرقمية في المنزل، وفي الشـارع، وفي أماكن العمل، وعبر شـبكات النقل، وفي غيرها، مما دفع المعلّقين للافتراض بأننا قد ولجنا عصر (البرمجيات في كل مكان -EveryWare) (Greenfield 2006). إذن، هذا هو عصر الحوسـبة واسـعة الانتشار في كل مكان، حيث يتمّ توزيع القدرة الحاسوبية وإتاحتها في كل مكان، مع العديد من الأجهزة الحاسـوبية التي تعمل لكل شـخص (Lucas et al. 2012)، فالحوسبة المنتشرة (Pervasive Computing) هـي إضافة للقدرة الحاسـوبية وإتاحة الوصول إلى شـبكات تكنولوجيا المعلومات والاتصالات للكائنات والبيئات اليومية الثابتة لجعلها تفاعلية و«ذكية» (Dourish 2001). وفي المقابل، فإن الحوسـبة في كل مكان (Ubiquitous Computing) هـي القدرة الحاسـوبية التي تتحرك مع الشـخص بغض النظر عن البيئـة؛ كالأجهزة التي يحملهـا النـاس مثل الهواتف الذكية وأجهزة الملاحة عبر الأقـمار الصناعية التي تعتمد على الوصول إلى شبكات تقنية المعلومات والاتصالات لتقديم الخدمات. وإذا كان شعار الحوسبة المنتشرة هو الحوسـبة «في كل شيء»، فإن شعار الحوسبة في كل مكان هو الحوسبة «في كل مكان»، مع إبراز الحوسبة المنتشرة لعمليات التباعد والتفرع (تضمين البرمجيات في مزيد من



الأجهزة) وإظهار الحوسبة في كل مكان لأوجه التقارب (جهاز رقمي واحد للقيام بمزيد من المهام) (Kitchin and Dodge 2011). وبينما يلزم الحوسبة المنتشرة أن تكون على إدراك موقعي/موضعي (Situationally Aware) لتنفيذها بنجاح، تتطلب الحوسبة في كل مكان الإدراك المستمر للسياق والموقع.

قابلية الفهرسة والقراءة الآلية (Indexical and Machine-Readable Identification):

على مدى نصف القرن الماضي، أصبحت الظواهر في العالم أكثر وأكثر تحديداً بشكل فريد ومن السهولة التقاطها كبيانات، بسبب التقنيات الرقمية وتقنيات العنونة ووضع العلامات الجديدة، مما زاد في درجة دقتها وعلائقيتها ومن ثم مستوى التفصيل في تحليلها، بالإضافة إلى تمكين ربط مجموعات بيانات أكثر (انظر الفصل الرابع). فعلى سبيل المثال، تتميّز كل الأجهزة الرقمية التي تساهم في الحوسبة المنتشرة والحوسبة في كل مكان بمعرّف فريد مُفَهْرس، بحيث يُعيّن رمز تعريف حصري لكل منها مما يتيح رصدها وتتبّعها على الدوام. وعلاوة على ذلك، فإن جميع عناصر إنترنت الأشياء (Internet of Things- IoT)، كوكبة واسعة من الأجهزة المتصلة بالشبكة، المحركات، وأجهزة الاستشعار، تنتج بيانات مفهرسة يمكن تمييزها بشكل فريد (Gershenfeld et al. 2004). وإضافة لذلك، فإن المعرفات الفريدة التي تم ربطها بالناس منذ زمن طويل، مثل جوازات السفر، والبطاقات الصحية، وسجلات التأمينات الاجتماعية، قد جرى تعزيزها من خلال أنظمة تحديد الهوية الرقمية والحيوية الجديدة التي تربطها بقواعد بيانات رقمية تتحقق من هويتهم (Dodge and Kitchin 2005). وهذه تشمل أسماء المستخدمين وكلمات السر، وبطاقات السحب المغناطيسية، والقارئات الرقمية لبصمة الأصابع، وقزحية العين، والحمض النووي. وفي كل حالة، تعد هذه المعرفات بمثابة المصافحة الرقمية، على الرغم من أنها تختلف في مستواها من حيث الدقة والثقة. فعلى سبيل المثال، يمكن لأسماء المستخدمين وكلمات المرور أن تُسرق أو تُزوّر بسهولة، في حين أن الحمض النووي هو أصيل وفريد لكل شخص ويحقق المعايير الأربعة الأولى من قائمة كلارك (Clarke 1994b) المتعلقة بالخصائص المرغوب فيها لتحديد هوية الكائن البشري بشكل فعّال وهي: شمولية التغطية / النطاق (يجب أن يمتلك الجميع الخاصية نفسها)؛ التفرّد (يجب أن يمتلك كل شخص قيمة واحدة فقط

لهذه الخاصية، ولا يتساوى فيها أي شخصين)؛ الدوام (يجب أن لا تتغير قيمة الخاصية بمرور الزمن، ولا تكون قابلة للتغيير أيضاً)؛ اللزوم (بمعنى أن تُعد الخاصية صفة طبيعية لا يمكن إزالتها ولا غنى عنها). وهذه الصفات تتوافق مع الصفات السبع الأخرى وهي: القدرة على التحصيل/ الجمع؛ والقدرة على التخزين؛ والتفرد؛ والدقة؛ والبساطة؛ والتكلفة والتناسب؛ والقبول.

وبالمثل، فإن رموز التحديد المكانية هي أنظمة تحدّد مواقع الأشخاص، والأماكن، والكائنات في جميع أنحاء العالم، وتشمل خطوط الطول والعرض، وإحداثيات الخرائط الشبكية، والوحدات الإدارية مثل الرموز البريدية (Zip Codes). وتختلف هذه الأنظمة في نطاقها ودقتها (درجة التفصيل)، وقد أصبحت أكثر دقة وتحديداً من أي وقت مضى. فعلى سبيل المثال، تُوسم البيانات الإدارية بالنسبة إلى الملكيات الفردية أو قطع الأراضي بدلاً من الأحياء؛ كما يسجّل أي جهاز يحوي نظام تحديد المواقع العالمي (Global Positioning System- GPS) الموقع في نطاق بضعة أمتار؛ وبدأت أنظمة الخرائط الوطنية بإسناد عدد كبير من الكائنات الناشئة بفعل بشري أو طبيعي في المخططات الخرائطية بمعرفات إقليمية فريدة. ويوجد معرفات فريدة أخرى شائعة على نحو متزايد للمنتجات (انظر الفصل الرابع)، والمعاملات (على سبيل المثال، طلبات الشراء)، والتفاعلات (على سبيل المثال، البريد الإلكتروني والمكالمات الهاتفية).

وفضلاً عن كونها أكثر قابلية للفهرسة والتبويب، أصبحت رموز تحديد الهوية قابلة للقراءة الآلية أيضاً على نحو متزايد (Dodge and Kitchin 2005). على سبيل المثال، يحتوي الرمز الشريطي (Barcode) على رمز تعريف يرتبط بنظام معلومات يحتفظ ببيانات وصفية إضافية تصف الكائن عند مسحه بقارئ ليزري (مثل نوع المنتج / النموذج، تاريخ ومكان الصنع، الأسعار، وما إلى ذلك). كما يمكن استشعار رقائق تحديد الهوية باستخدام موجات الراديو (RFID) عن بُعد من قبل مستجيب راديوي والتعرف عليها. ويمكن كذلك التعرف على المركبات بواسطة برنامج التعرف التلقائي على أرقام لوحاتها، وقد تمّ تطوير تقنيات خوارزمية جديدة للتعرف وتحديد هوية الوجوه وطريقة المشي (كيف يمشي الناس). وهكذا أصبحت عناصر / كائنات العالم مفتوحة لجمع والتقاط بياناتها وتحديد

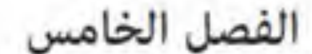


هويتها تلقائياً، مما يوسّع بشكل كبير نطاق البيانات المنتجة ومدى دقتها بالنسبة إلى نظام ما، مما قدّم أشـكالاً جديدة من التنظيم والحوكمة (Dodge and Kitchin 2007a) وعزّز عمل الأنظمة بطرق أكثر آلية (Kitchin and Dodge 2011). لذا فإن الاتجاه السائد هو المزيد والمزيد من التحديد للظواهر والممارسـات لتصبح معرّفة بشكل فريد وقابلة للقراءة آلياً، مما يزيد الدقة، والعلائقية، والأتمتة.

تخزين البيانات (Data Storage):

كان هنـاك تحوّل كبير، عـلى مدى العقدين الماضيين، في القدرة عـلى التخزين الطويل الأمد لكميات هائلة من البيانات، ففي البداية تمّ تخزين البيانات الرقمية في صورة تناظرية باسـتخدام البطاقات المثقوبة وذلك قبل أن يتم الانتقال إلى الشريط المغناطيسي والأقراص الصلبة. وقد اسـتكملت هذه الوسائل باستخدام وسـائط التخزين الرقمية المتنقلة والتي تُعرف باسـم الأقراص المرنة (Floppy Disks)، وتـلا ذلك ظهور الأقراص المدمجة (CDs)، وأقراص الفيديو الرقمية (DVDs)، ومحركات الأقراص الخارجية والذاكرة المتنقلة (Flash/ External Drives). وقـد دفعت تقنيات الأقراص الصلبة (Hard-disk Technologies)، عـلى وجه الخصوص، بالقدرة على تخزين البيانات الرقمية بشـكل كبـير، وذلك مع النمو الكبـير للسـعة التخزينية المقدرة بالخانة الثنائية لكل بوصـة مربعة (بت/إنش2)، في حين انخفضت التكلفة لكل غيغابايت بشكل كبير وبعامل يقدّر بعشرة ملايين إلى واحد من عام 1956 م إلى 2000م (Gilheany 2000) واسـتمرت الأسعار في الهبوط، مما جعل التخزين الرقمي أرخص تكلفة بكثير وأقل مساحة من الأشكال التناظرية (على سبيل المثال، الورق أو الأشرطة الفلمية) (Grochowski and Halem 2003). فالآن، أصبحت التكاليف رخيصة جداً وبالشـكل الـذي يُمكن من تخزين كميـات هائلة من البيانـات بمقابل بخس، ودون الحاجة تقريباً لحذف الملفات القديمة. ونتيجة لذلك، فإن المُعاملات التي تجري اليوم (على سـبيل المثال، رسالة أرسـلت إلى قائمة البريد الإلكتروني أو صفحة موقع التواصل الاجتماعي الفيسبوك، أو المدفوعات إلى متجر باستخدام بطاقة الائتمان، أو المحادثة مع أحد الأصدقاء على الهاتف) من الممكن تسجيلها وحفظها بشكل جيد إلى ما بعد وفاة الشخص المعني بها، مع إمكانية الإشارة إليها وتحليلها في أي وقت في المستقبل.





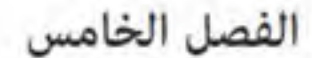
ونتيجة لذلك، وكما نوقش في الفصل السابق، كان النمو في حجم البيانات المخزنة بالأرقام المطلقة والنسبية استثنائياً، وخاصة منذ العام2000م. فعلى سبيل المثال، قدّر Hilbert and López (2009) أن خمسة وعشرين في المائة فقط من البيانات تم تخزينها في شكل رقمي في عام 2000، مع الاحتفاظ بما تبقّى بأشكال تناظرية مثل الكتب، والمجلات، والصور، والأشرطة الممغنطة. وبحلول عام 2007 م، شكّلت البيانات الرقمية ما نسبته أربع وتسعون في المائة من مجموع البيانات المخزنة، ومنذ ذلك الحين تواصل النمو للحصة النسبية من البيانات الرقمية، خاصة مع تطور التخزين والخدمات الموزعة من خلال الحوسبة السحابية ومراكز البيانات. وتأخذ الحوسبة السحابية أحد شكلين يعملان في الغالب بصورة تعاونية: سُحُب المرافق الخدمية (Utility Cloud)، وسُحُب البيانات التخزينية (Data Clouds) (Farber et al. 2011). فالسُحب الخدمية توفّر القدرات والإمكانات المتعلقة بتقنية المعلومات كخدمات حسب الطلب حيث يُمكن الوصول إليها عن طريق شبكة الإنترنت وذلك بشكل مستقل عن موقع طالب الخدمة. ويشمل ذلك «البنية التحتية كخدمة» (Infrastructure as a Service- IaaS) مثل التخزين والخوادم والشبكات، «المنصة كخدمة» (Platform as a Service- PaaS) وتضم بيئة التنفيذ لتطوير تطبيقات مخصصة وقواعد البيانات، و«البرمجيات كخدمة» (Software as a Sservice- SaaS) والتي تمكن المستخدمين من الوصول إلى التطبيقات ومعالجة البيانات عن بُعد (Farber et al. 2011; Hancke et al. 2012). في حين تُمكّن سحب البيانات من ربط، تخزين، ومعالجة كميات هائلة من البيانات التي قد تنتج عبر المؤسسة عن بُعد، بالاعتماد على القدرة الحاسوبية لمئات من الأجهزة، وتحليلها عن طريق سُحب المرافق الخدمية (Farber et al. 2011). ومن ثم، يمكن للأفراد والشركات الاستفادة من التخزين والقدرة الحاسوبية دون الحاجة إلى القيام باستثمارات رأسمالية كبيرة، فضلاً عن كونها قادرة على الاستفادة من هذه الموارد من أي مكان يتوافر فيه اتصال بالإنترنت (Bryant et al. 2008).

وبالمثل، كان هناك تحوّل في كيفية تنسيق البيانات وهيكلتها عند تخزينها، وبسبب الزيادة الناشئة في حجم، وسرعة، وتنوّع البيانات، فقد عانت قواعد البيانات العلائقية للبقاء ومواجهة الموقف. فقد صُمّمت لبيانات على درجة عالية من التنظيم، والسلامة،



والمنظمة، والثابتة. وبدلاً عنها، ظهر شــكل جديد من قواعد البيانات هو قواعد البيانات غــير الهيكلية (NoSQL)، مــما يتيح تخزين البيانات الكبيرة. وتميــل قواعد البيانات غير الهيكلية إلى أن تكون أقل تعقيداً في الشــكل من قواعــد البيانات العلائقية، ولكن يمكنها التعامل مع كل من البيانات المنظمة (Structured Data) وغير المنظمة (Unstructured Data)، وهــي قابلة للتوســع بسرعة، ويمكنهــا التعامل فعلاً مع العديــد من مجموعات بيانات سريعة التغير وهائلة في الحجم، وتمتاز بالمرونة في كيفية تنظيمها، مما يتيح هياكل بيانــات مفصّلــة بحســب الطلــب (Driscoll 2012; Marz and Warren 2012). كما تميــل قواعد البيانــات غير الهيكلية إلى تخزين البيانات الابتدائية الخام بدلاً من المشــتقة التي يتم هيكلتها وفقاً لعلاقة محددة ســلفاً (Driscoll 2012). فعلى سبيل المثال، تخزّن قواعــد البيانــات غير الهيكلية معلومــات خام عن زيارة صفحة الويب (المشــاهدة) بدلاً من تخزين تعداد المشــاهدات للصفحات المشــتقة منها، وهي إســتراتيجية أكثر قوة من حيث أنها تُقلّل من الأخطاء المحتملة في قواعد البيانات، ولكن هذه الإســتراتيجية لا يمكن تطبيقها مع قواعد البيانات العلائقية بســبب النمو السريع والحجم الناتج من ذلك النمو والــذي لا يمكن التعامل معه من خلالهــا (Marz and Warren 2012). ويتم في العادة توزيــع البيانات وتكرارها عبر العديد من الأجهزة في قواعد البيانات غير الهيكلية، بدلاً من مركزيتها في مكان واحد (ومن ثم حل مشكلة مجموعات البيانات الكبيرة جداً كونها كبيرة جداً عن أن يجري احتواؤها على جهاز واحد)، وتُقسّم طلبات الاستعلام ويجري تنفيذها في الوقت نفسه على عدد من الأجهزة جنباً إلى جنب ثم يُعاد تجميع النتائج (ومن ثم تسريع العمليــات الحســابية) (Dumbill 2012). وتتلخص هذه اللإســتراتيجية من خلال تقنية التعيين والاختزال (ReduceMap) التي تمّ تطويرها من قبل شركة جوجل حيث يتم تعيين البيانات والاســتعلامات على عــدة خادمات ومن ثمّ جمع النتائــج الجزئية واختزالها معاً (Dumbill 2012). ويُعد هادوب (Hadoop)، الذي تمّ وضعه بداية من قبل شركة ياهو (Yahoo)، مثالاً على قواعد البيانات غير الهيكلية ومفتوحة المصدر وشــائعة الاستخدام في تطبيــق نموذج التعيين والاختزال. فالحل البديل لمثل هذا النهج يتمثّل باســتخدام نموذج أكــثر محدودية للبيانات، ودرجة أقل مــن العلائقية في مجموعة البيانات، وقدرة أقل على الاستكشاف باستخدام لغة الاستعلام العلائقية (SQL) (Driscoll 2012). وهذا يعني أن





قواعد البيانات غير الهيكلية مثالية للتخزين الفعّال والاسترجاع السريع لكميات هائلة من البيانــات، ولكنها أقل فائدة لفحص العلاقات بين عنــاصر البيانات (مركز تحالف البيانات المفتوحة Open Data Center Alliance 2012

مصادر البيانات الكبيرة (Sources of Big Data):

وفّر التوســع الجذري والتكامل بين الحوســبة، والشــبكات، والأجهزة الرقمية، وتخزين البيانــات منصة قوية لانطلاقة البيانات الكبيرة، وإضافة لذلك الوســيلة التي يتم بها إنتاج، ومعالجة، ومشاركة، وتحليل البيانات الكبيرة. ويمكن تصنيف هذا الإنتاج للبيانات في ثلاث فئــات رئيســية. البيانات الموجّهــة (Directed Data) والتي يتمّ إنتاجها بوســائل مراقبة تقليدية، حيث تتركّز أنظار التقنية على شخص أو مكان من قبل مشغّل بشري. أما في حالة البيانات الآلية (Automated Data)، فتنتج البيانات عن وظيفة كامنة في الجهاز أو النظام بطريقة تلقائية، في حين يتم تداول البيانات الطوعية (Volunteered Data) وإهداؤها من قبل الناس للنظام.

البيانات الموجّهة (Directed Data):

توجــد المراقبة المنظمة والهيكليــة حيثما وُجدت مجموعة من النــاس الذين يشرفون على أشــخاص آخرين كالمســؤولين عن إنفاذ القانون، والمعلمين، والأطباء، ومسؤولي الرعاية الاجتماعية، والبيروقراطيين، ورؤســاء العمل والذين يعملون على مراقبة الآخرين شــخصياً كالمواطنين، والتلاميذ، والمرضى، والعمال أو من خلال العدسات التقنية كالمسوح مثل التعداد الســكاني، ونماذج البيانــات الحكومية، وإيصالات الضرائب، وعمليــات التفتيش، وكاميرات مراقبــة الدوائــر التلفزيونية المغلقــة (CCTV)، حيث كانت المراقبــة المنظمة والهيكلية منذ فترة طويلة ســمة من سمات المجتمعات، ومكوّناً أساسياً من مكونات الدولة وحوكمة الشركات (Lyon 2007). فهذا النموذج للتحكم (التداخل المنطقي، الأدوات، المؤسســات، وأدوار وإجــراءات الحوكمــة) يُتيــح مركزية التنظيــم والتحكم عبر مجموعة واســعة من المجالات، ويساعد في الحفاظ على النظام، وإنتاج حكومة جيدة، وإدارة فعّالة، وعمل مربح، ومجتمعات مســتدامة ومســتقرة، من خلال المحاسبة الفعّالة للأشــخاص والتهذيب الذاتي

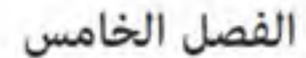


أيضـاً (بمعنى، تعديل الناس سـلوكهم ليتوافق مع التوقعات والقواعد). وتُسـتكمل هذه النظـم بطرق تحديد هوية الأفراد، مثل بصمات الأصابع، والصور، وأرقام التأمين الوطني أو الضمان الاجتماعي، وجوازات السـفر ورخص القيادة، التي تجعل من السـهل تتبع واقتفاء أثر الناس، كما تمّ توسـيع نطاق المراقبة ليشـمل الحيوانات والبيئات، واستخدام الكائنات وحركتها.

وحتـى وقت قريـب، كانت كل عمليـات المراقبة جزئيـة وتناظريـة بطبيعتها وتنتج مسـتويات مختلفة من البيانات المسـجّلة، بدءاً من الملاحظات غير المسجّلة إلى السجلات التفصيلية أو التسـجيل المسـتمر، ويجري تطبيقها أحياناً على عينات وفي حالات قليلة على المجتمـع بأكمله. وكانت البيانات المسـجّلة انتقائية، تُجمع من خـلال التقارير أو الإبلاغ الذاتي على الورق أو على شريط مغناطيسي أو فيلم، في مكان وزمان معيّنين. وتميل البيانات إلى أن تكون منظمة للغاية في طبيعتها، ومقتصرة على مجموعة من حقول البيانات المحددة مسـبقاً. وكانت السـجلات ضخمة ومكلفة في التخزين، مما يعني عدم الاحتفاظ الطويل الأمـد بها إلا لكمية محدودة من البيانـات، إضافة إلى كون الكثير منها إما عبارة عن عينة أو مشتقة. وحتى في الحالات التي قد تكون المراقبة فيها مستمرة، مثل الدوائر التلفزيونية المغلقة، فقد تمتدّ اللقطات نفسها على فترات زمنية طويلة تتعدى قدرة الشريط التسجيلي على تغطيتها ويجري في العموم إعادة التسـجيل عليها بعد فترة معينة (غالباً بعد سـبعة أيام أو شـهر). كما كان تحليل البيانات الناتجة من مختلف أشـكال المراقبة المسـتخدمة يتمّ بواسـطة مفسّر بشري إلا إذا جرى تحويلهـا رقمياً في وقت لاحق. وفي حين أن كميات البيانات التي تنتجها هذه المراقبة يمكن أن تكون ضخمة، وربما تستوفي واحدة أو اثنتين من الخصائص المرتبطة بالبيانات الكبيرة مثل قابلية الفهرسـة والشمولية والعلائقية أو السرعة، إلا أنه لا يمكنها بأي حال أن تستوفي مجموعة الصفات المرتبطة بالبيانات الكبيرة كاملة.

وفي حين لا تزال العديد من أنظمة المراقبة والحوكمة مستمرة في كونها جزئية وتناظرية، كان هناك تحرّك في الآونة الأخيرة لاسـتبدالها والتوسع فيها بنظائر رقمية مكافئة لها والتي تنتـج الآن بيانات كبيرة. فعلى سـبيل المثال، يجري اسـتبدال أنظمة الدوائـر التلفزيونية المغلقة التماثلية الثابتة بأخرى رقمية متحركة حيث يمكن تغيير مجال الرؤية، والتقليل من





المحددات على التخزين، وتعزيز التفسير البشري بتحليلات خوارزمية حسابية، مثل التعرف على الوجه أو طريقة المشي، كما يمكن ربط مثل هذه البيانات مع قواعد بيانات أخرى. وهنا، يوجّه تركيز الكاميرا، وعمليات التحليل والتفسير، من قبل مشغّل بشري، وبمساعدة من البرمجيات. على سبيل المثال، من الممكن مراقبة وضبط العديد من الكاميرات في غرفة عمليات الشرطة وربطها بسجلّ الحوادث الحية من أجل الاستجابة بكفاءة وتوجيه الموارد المناسبة لمواقع معينة بشكل استباقي. وبالمثل، في مراقبة الهجرة والجوازات، حيث يجري آنياً جمع معلومات الركاب التفصيلية وتدقيقها مع مختلف قواعد البيانات الأمنية والحدودية بشكل آني أو لحظي (Real-Time)، وإنتاج بيانات جديدة مثل المسح الضوئي للصور الشخصية وبصمات الأصابع أو قزحية العين. وتستخدم العديد من الحكومات المحلية أنظمة إدارية لتسجيل مشاركات المواطنين مع خدماتها آنياً ورصد إن كان تعامل الموظفين مع أي قضايا قد تمّ ضمن أطر زمنية محددة، مع مراقبة مديري الإشراف للنظام لإعادة توجيه الموارد بحسب الحاجة.

وتتكوّن بعض الأنواع الأخرى الأقل انتظاماً في إنتاج البيانات الكبيرة الموجّهة من: التصوير الجوي الرقمي عن طريق الطائرات أو الطائرات بدون طيار، أو الفيديو المكاني، أو الليدار (الرادار بتقنية الليزر) لكشف الضوء وتحديد مدى البُعد، أو الحرارية أو غيرها من أنواع الأشعة الكهرومغناطيسية للبيئات التي تُمكّن من الرسم المتحرك للخرائط الثنائية والثلاثية الأبعاد للمناظر الطبيعية المفهرسة مكانياً بشكل آني أو لحظي (Real-Time) من خلال نظام تحديد المواقع العالمي (GPS). وتنتج مثل هذه الفيديوهات والمسوح بالأشعة كميات هائلة من البيانات المتنوعة عالية السرعة. فعلى سبيل المثال، يمكن لماسح الليدار توليد مليون نقطة بيانية، مع دقة بيانية تصل إلى بضعة سنتيمترات، مع إمكانية إجراء أكثر من مسح خلال الثانية الواحدة. إن وضع ماسح ليدري في مركبة وقيادتها في الأنحاء، أو الطيران به فوق منطقة معينة ينتج تمثيلا ثلاثي الأبعاد مفصّلاً للغاية للتضاريس الطبيعية لتلك المنطقة.



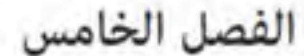
البيانات الآلية (Automated Data):

في حين وسّعت البيانات الموجّهة أنظمة المراقبة والحوكمة التقليدية، فقد حوّل إنتاج البيانات الآلية بشكل جذري كيفية إنتاج البيانات وتوظيف استخداماتها، ومكّن من إنشاء أنظمة جديدة للتنظيم والإدارة التلقائية (Dodge and Kitchin 2007a, Kitchin and Dodge 2011؛ انظر الفصل العاشر). فمن خلال مجموعة من التقنيات الرقمية يتمّ إنشاء البيانات تلقائياً بواسطة مجموعة من النظم الآلية وبرقابة بشرية قليلة. وعلاوة على ذلك، يتمّ في الغالب معالجة وتحليل مثل هذه البيانات تلقائياً وبشكل مستقل، إذ يجري العمل عليها بواسطة خوارزميات برمجية لتحقيق نتيجة معينة تهدف إلى تنظيم ظاهرة ما. وهناك طرق عديدة لإنتاج البيانات الآلية، بعضها قد يكون إنتاجه ثانوياً عن النظام وليس من صميم عمله.

المراقبة الآلية (Automated Surveillance):

كما أصبحت تقنيات المراقبة رقمية بطبيعتها ومرتبطة بعضها ببعض، فقد أصبح من الممكن أيضاً أتمتة جوانب متعددة لنظام المراقبة، وإضافة تقنيات جديدة، لتتبع أكثر فعالية وكفاءة في مراقبة وتتبع الاستخدام للأنظمة والأماكن المختلفة. وتعدّ قراءة العدادات الذكية مثالاً على النموذج اليدوي من المراقبة التي غدت آلية بشكل متزايد. هنا، يتم استخدام تقنية قراءة العداد آلياً (Automatic Meter Reading- AMR) لمراقبة استخدام الآلة والتواصل معها دون الحاجة إلى القراءة اليدوية التقليدية (Hancke et al. 2013)، وعلاوة على ذلك، يمكنها القيام بهذه المهام بشكل متواصل مما يمكّن المورّد من تتبع الاستخدام بشكل آني أو لحظي، وهذا له فائدة في مطابقة الطلب مع العرض وفي العثور على أخطاء / تسريبات في النظام. كما أنها توفّر وسيلة للقيام بالفوترة الآلية، مما يحدّ من نفقات التوظيف. وفي كثير من الحالات، من الممكن للمستهلك أيضاً أن يراقب استهلاكه الخاص لخدمة معينة، مثل الكهرباء أو الغاز أو المياه، فعلى سبيل المثال، يمكن استخدام تطبيق للهاتف الذكي للاتصال بعداد عن بعد باستخدام تقنية اتصال المجال القريب على سبيل المثال، البلوتوث، ومن ثم تغيير الاستهلاك تبعاً لذلك.





وبالمثـل، تعزّز أنظمـة إدارة المرور من القدرة على تنظيم وضبط السـائقين من خلال الاستفادة من تقنية التعرف الآلي على أرقام لوحات المركبات (Automatic Number Plate Recognition- ANPR) (Dodge and Kitchin 2007a). فهنـا، تمسـح الخوارزميـات بيانـات الصور مـن الكاميرات الرقمية للكشـف عن لوحات الترخيـص. ومن جهة، يمكن استخدام هذه المعلومات لتتبع المركبات خلال عبورها على طرقات المدينة وتقديم مدخلات إلى نظم النقل الذكية (Intelligent Transportation Systems- ITS)، ومن جهة أخرى، يمكـن إحالة هذه البيانـات إلى قاعدة بيانات مالكي المركبـات لإدارة الغرامات والعقوبات المتعلقة بالمخالفات المرورية. فعلى سـبيل المثال، بالنسبة للمثال السابق، يتم فحص لوحة الترخيص لجميع المركبات الداخلة إلى منطقة الزحام المروري في لندن ومطابقتها مع قاعدة بيانات تلك المركبات التي سـددّت رسـوم الازدحام، أما تلك التي لم تسدّد خلال فترة أربع وعشرين سـاعة فيتم تغريم مالكيها تلقائياً من خلال عمليـة الإدارة الآلية (يتمتع النظام باسـتقلالية في إصدار الغرامات دون الحاجة لرقابة بشرية). ويمكن للنظام بالمثل استخدام كاميرات السرعة لإصدار قسائم تجاوز السرعة للسائقين المخالفين.

وفي حالات أخرى، تمّ تسـهيل المراقبة الآلية من خلال اسـتخدام رموز الهوية المقروءة آلياً لتسـجيل ماهية الأنشـطة المريبة/ المجهولة على شبكة المراقبة. فعلى سبيل المثال، تمّ اسـتبدال التذاكـر الورقية التي يتزايد تبادلها في السـفر على نظام النقـل من غير حامليها بالبطاقـات الذكية، التي ينبغـي التدقيق عليها لدخول وخروج المحطات ووسـائل النقل. ويتمّ إنشاء سجل عند نقاط المسح ومن ثم يمكن تتبع كل حركات حاملي البطاقات الذكية. وفي حالـة نظام النقـل في لندن، فإن اسـتخدام بطاقات أويسـتر (Oyster) يعني أنه من الممكن تتبع الرحلات الفردية عبر نظام الحافلات والسـكك الحديدية المسـتخدمة من قبل مليـوني راكـب يوميـاً (Batty et al. 2012). وبالمثل، من الممكـن الآن المراقبة التلقائية لأشـكال متنوعة من الخدمـات العامة المقدمة التي كان يجري رصدهـا عرضاً، مثل جمع النفايات، من خلال استخدام رقائق تحديد الهوية باستخدام موجات الراديو (RFID) التي تثبّت على حاويات النفايات ويتم مسـحها عن بُعد عند نقاط الجمع. وأصبح من الممكن قياس كميات التخلص من النفايات باسـتخدام مقاييس مرفقة بمجموعة مركبات وتحميل الأسر بصورة فردية تكلفة التخلص من النفايات على أسـاس الحجم بدلاً من الرسم الموحّد

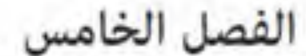


دون تمييــز. كما أصبــح من الممكن أيضــاً الحصول على فهم مفصّل لأنمــاط التخلص من النفايات داخل منطقة ما وإعادة تخطيط طرق وأوقات الجمع.

الأجهزة الرقمية (Digital Devices):

كما ســبق الإشــارة إليه، كان هناك نمو هائل في عدد الأجهزة الرقمية المستخدمة حول العالم والتي تســاعد على تسهيل، وتعزيز، وإدارة العديد من جوانب الحياة اليومية. فهذه الأجهــزة لا تقوم بمجرد معالجة البيانات الرقمية والتعليمات فحســب، ولكنها تنتج بيانات رقمية بشكل نشط وكوظيفة أساسية لها أو ثانوية، أو الاثنين على حدّ سواء. ومن الأمثلة على الأجهــزة التي تنتج بيانات رقمية كوظيفة أساســية: الكاميرات، وأشرطة الفيديو، والمراقيب ووحدات تحديد المواقع (GPS)، وأشكال مختلفة من المعدات الطبية مثل كاميرات المنظار، والماســحات بالموجات فوق الصوتية، وموازين الحرارة الرقمية، إلخ. أما أمثلة الأجهزة التي تنتج بيانات رقمية كوظيفة ثانوية، فتشــمل الهواتف النقالة ومســتقبلات البث التلفزيوني عــبر الكابل أو الأقــمار الصناعية، والتــي تنتج بيانــات متعلقة بكيفية اســتخدامها (على ســبيل المثال، الوقت، المكان، الشــخص المتصل / القناة المشــاهدة)، مــع البيانات المتصلة بهــا والتي جرى جمعها في ســجلات يتم إرســالها إلى أطراف ثالثة (مثــل مصنعي الأجهزة ومقدمــي الخدمات). أما الأمثلة على الأجهزة التي تنتج البيانات الأساســية والثانوية معاً فتشــمل الهواتف الذكية التي يمكن استخدامها لتسجيل الملاحظات، والتقاط الصور، وكتابة البريد الإلكتروني والتعليقات على وســائل التواصل الاجتماعي، ...إلخ، بالإضافة إلى تســجيل اســتخدام مختلف التطبيقات. وتشكّل هذه الأجهزة ســلالة جديدة من الكائنات، تُسمى لوجكتــس (Logjects) (Kitchin and Dodge 2011) – كائنــات ترصــد وتتبــع تاريــخ اســتخدامها الخاص. وفي كثير من الحالات، يكون من الممكن تســجيل موقع هذه الأجهزة. فالهاتــف الجوال والهواتــف الذكية يمكن تتبّعها من الفضاء عن طريق التثليث (حســاب المثلثات) عبر هوائيات الهاتف، وأيضاً عن طريق مســتقبلات تحديد المواقع المدمجة فيها، بالإضافة إلى مراقبة السرعة والاتجاه من خلال المدوار (Gyroscopes)، ومقاييس التســارع (Accelerometers)، والبوصلات (Compasses) على شــكل أقراص أســطوانية، أو أجهزة ملاحــة بالأقــمار الصناعية أو أي أجهزة أخــرى. وعلاوة على ذلك، فــإن العديد من هذه





الأجهزة يمكن أن تتفاعل بشكل مستقل بعضها مع بعض وتقوم بتبادل البيانات. ومع تكاثر الأجهزة المعقــدة لكافة المهام وفي كل الأماكن، يتمّ إنتاج المزيد والمزيد من البيانات الغنية بالبيانات الوصفية القابلة للفهرســة، مما يتيح تحليلاً أكثر تطوراً واســتحداث إستراتيجيات ومنتجات جديدة قائمة على البيانات.

البيانات المستشعرة (Sensed Data):

تعد المستشــعرات والمحركات من أنواع الأجهزة الرقمية. وهي ما يمكن أن تكون ضمناً أو جزءاً من هياكل مختلفة لقياس مخرجات معينة مثل مستويات الضوء، والرطوبة، ودرجة الحرارة، والغاز، والمواد الكيميائية، والمقاومة الكهربائية، والصوتيات، وضغط الهواء، والحركة، والسرعة، ...إلخ. وتُعد أجهزة الاستشعار والمحركات صغيرة، وغير مكلفة في العموم، ويمكنها إنتاج تيار مســتمر من البيانات، ويمكن أن تكون سلبية وتقرئ بواسطة الماسحات الضوئية، أو يمكن أن تكون نشــطة، بحيث تبثّ البيانات على فترات منتظمة عبر شــبكات الحاسب المحليــة أو الواســعة، أو قد يكون لديها قدرة الاتصال من خــلال حقل الاتصالات القريب (NFC) والذي يتيح الاتصالات في اتجاهين (Hancke et al. 2013). وبوضعها على جسر، يمكن للمستشعرات أن تقيس وتتواصل لحساب معدلات التآكل (عن طريق قياس المقاومة الكهربائية) ومقدار الضغط على مواد البناء باســتخدام أجهزة استشعار الانبعاثات الصوتية (للكشف عن انتشار الموجات الصوتية) وأجهزة استشعار التخصّر المغناطيسي (للكشف عن تغيرات في الحث المغناطيسي)، ومن ثم يقلل ذلك من الحاجة إلى التفتيش اليدوي المكلف (Hancke et al. 2013). وبالمثــل، يمكن للمستشــعرات داخل نظــام المياه قياس نوعية المياه، وضغطها، ومقدار تدفقها، وتمكين الإدارة الآنية وصيانة الأنابيب لحظة بلحظة. وعلى المركبات، يمكن للمستشــعرات رصد حجم العمل، والإجهــاد، والتضاريس. فشركة هيتاشي تستخدم هذه المجسات لرصد مركباتها الثقيلة المعقّدة والمكلفة، وكذلك تفعل الشيء نفسه شركــة النقل السريع (UPS) مع أســطولها مــن عربات النقل والشــاحنات بحيث توظّفها لتشغيل خدماتها بكفاءة وبصورة اســتباقية (Plumridge 2012; Mayer-Schonberger and Cukier 2013). ومــن خــلال ربط رقائق تحديد الهوية باســتخدام موجات الراديو (RFID) بالمنتجــات أصبــح من الممكن رصــد وتتبع حركة الوحــدات الفردية من المصنع



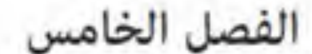
أو المزرعة إلى المستهلك، والحدّ من السرقة والهدر. وبالمثل، تحتوي بطاقة الضريبة على شريحة تحديد الهوية باستخدام موجات الراديو التي تتواصل مع قارئ ترددات راديوي على حواجز تحصيل الرسوم، مما يؤدي إلى سرعة أعلى في حركة المركبات والدفع الآلي، بالإضافة إلى حساب معدل حركة المركبات أو عدد مواقف السيارات المتاحة لإدارة حركة المرور. كما تثبّت رقائق تحديد الهوية باستخدام موجات الراديو على الحافلات والقطارات والتي تتواصل مع قارئ الموجات الراديوية على طول مساراتها مما يجعل من الممكن تتبع موقع المركبات في أي لحظة آنياً.

كما يمكن إنشاء شبكة استشعار كثيفة من خلال توزيع العديد من أجهزة الاستشعار وربط البيانات التي تنتجها، مما يتيح رصد أوضاع مختلفة عبر نظام أو مكان ما. وقد توالت الأمثلة على هذه الشبكات وأجهزة الاستشعار في مدينة سانتاندر (Santander) في إسبانيا كجزء من مشروع المدينة الذكية (SmartSantander)، وفيه 12000 من أجهزة الاستشعار، وتتكون من خمسة أنواع، تمّ توزيعها بشكل إستراتيجي في جميع أنحاء المدينة لقياس الضوضاء، والحرارة، ومستويات الإضاءة المحيطة، وتركيز غاز أول أكسيد الكربون، والأماكن المتوافرة لوقوف السيارات وركنها (Bielsa 2013). وترتبط أجهزة الاستشعار معاً من خلال الاتصالات اللاسلكية وتوظيف البرمجة عبر الأثير (OTAP)، مما مكّن من برمجتها وتحديث برمجياتها عن بُعد (Bielsa 2013). كما يجري تطوير وادي بلان آي تي (PlanIT) في البرتغال، وهو في الوقت الحاضر تحت الإنشاء ومصمّم لإيواء 225 ألف نسمة، ويهدف إلى إنشاء بيئة قائمة على أكثر من 100 مليون جهاز استشعار مدمج فيها والتي سوف تنتج بيانات لرصد مجموعة متنوعة من البنى التحتية والبيئات (Marchetti 2012). وهكذا، يُنظر إلى المستشعرات بوصفها عنصراً أساسياً في التطورات التي تشهدها المدينة الذكية.

بيانات المسح الضوئي (Scan Data):

بالإضافة إلى الأجهزة الرقمية وأجهزة الاستشعار التي تنتج البيانات بشكل نشط والتي هي جزء من شبكة إنترنت الأشياء، من الممكن رصد وتتبّع الكائنات الأخرى رقمياً من خلال رموز الهوية المقروءة آلياً. وقد تجاوز استخدام الرمز الشريطي (Barcode) قطاع البيع بالتجزئة ليشمل السلع المصنّعة والاستمارات الرسمية، كما وضعت العديد من المنظمات





الكـبرى والقطاعات الصناعية نموذجهـا الخاص من الرمز الشريطـي ومعاييرها لتخصيص الأرقام (Dodge and Kitchin 2005). وتُمثّل البيانات في الرموز الشريطية ثنائية الأبعاد أفقياً ورأسـياً بمجموعة من النقاط المرقطة ويمكن اسـتخدامها لتحديد كائن بشكل فريد. وتسـتخدم بعض خدمـات البريد هذه الرمـوز الشريطية الثنائية الأبعـاد لتمكين المتابعة والتعقيب للرسـائل والطرود البريدية بشكل فردي. وبالمثل، تحتوي الشرائط المغناطيسية المثبتة على بطاقات الائتمان والولاء وبطاقات السماح بالعبور من خلال الأبواب على معرّف فريد يختصّ بصاحب البطاقة التي يتم قراءتها عند تمريرها على القارئ. وقد جرى استبدالها لاحقاً برقائق إلكترونية يمكن التحقق من البيانات التفصيلية فيها باستخدام الأرقام السريّة الشـخصية لمستخدميها. كما يتمّ إنتاج البيانات الخاصة بالاستعمال، في كل مرة يجري فيها مسح الرمز الشريطي أو الشريط المغناطيسي أو الرقاقة الإلكترونية، متضمناً ذلك أين ومتى تمّ الاسـتخدام ولأي غرض. وبالنظر إلى أن الغرض من هذا المسـح في كثير من الحالات هو لشراء السلع، فهي توفّر وسائل مهمة لإنتاج معلومات عالية الدقة والتفصيل للصفقات المبرمة.

البيانات التفاعلية (Interaction Data):

عـلى الرغم من أن بعض مسـتخدمي الإنترنت ينظرون إليها باعتبارها وسـطاً مجهول المعالم إلى حدّ ما، إلا أن كل تفاعلات المستخدمين الجارية عبر شبكات المعلومات والاتصالات تُنتج بيانات تفاعلية. وحقيقة، فإن إنتاج البيانات هو قائم بطبيعته على العديد من أشكال الاتصال والاسـتخدام في شـبكات تقنية المعلومات والاتصالات. على سـبيل المثال، يتعقّب مزودو خدمات الإنترنت المواقع التي تتصل بها حسـابات المسـتخدمين الفردية والأجهزة، وتسـتخدم العديد من المواقع، وخاصة مواقع الشركات، سـجلّ التتبـع وتعريف الارتباط للمتصفح (كوكيز Cookies) لإنتاج بيانات النقر (اختيارات المسـتخدم)، لتتبّع كيفية تنقّل المسـتخدم خلال الموقع على شبكة الإنترنت وما الميزات التي جرى اختيارها. وبشكل عام، تستخدم مثل هذه البيانات لمراقبة كيفية أداء الموقع الإلكتروني، وإجراء أبحاث عن السوق وتتبّع مسـار إنتاجيـة الموظف، مع إمكانية بيع هـذه البيانات إلى أطـراف أخرى. ويتمّ إنشـاء بيانات النقر هذه بصورة تلقائية ما لم يُعدّ المسـتخدم تطبيق المتصفح لعدم قبول ملفات تعريف الارتباط (وهذا قد يضعف اسـتخدام الموقع). وبالمثل، يتم تخصيص رموز

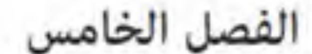


تعريفيــة فريدة في جميع الصفقات المالية القائمة عــلى تكنولوجيا المعلومات والاتصالات، سواء الخدمات المصرفية الشخصية الإلكترونية، أو عمليات سحب أموال من أجهزة الصراف الآلي، أو عمليات المتاجرة في الأسهم، ويجري تخزين بيانات هذه المعاملات تباعاً. وتحتوي جميع رسائل البريد الإلكتروني على رؤوس في مقدمة حزم بياناتها والتي تقوم بتحديد البريد الإلكتروني بشكل فريد، وتحتوي على معلومات تشمل المرسل والمتلقي، والخوادم والأجهزة المســتخدمة. وعلى المنوال نفســه، يتم تمييز جميع المكالمات الهاتفيــة بمعلومات المتلقي للمكالمــة، ووقت إجرائها ومدة بقائها، وما إذا كان الهاتف محمولاً، والموقع. ونتيجة لذلك يتم إنشــاء كميات هائلة من البيانات بشكل روتيني بشأن التفاعلات الحاصلة عبر شبكات تكنولوجيا المعلومات والاتصالات.

البيانات التطوعية (Volunteered Data):

على النقيض من المراقبة الموجهة إلى الأشخاص أو الأشياء من قبل الأفراد والهيئات، أو تلــك التي يتم التقاطها تلقائيا باعتبارها ميزة متأصلــة لجهاز أو نظام ما، فإن كثيراً من البيانات الكبيرة يتم إنتاجها من قبل الناس بشــكل تطوعي فعّال. ففي مثل هذه الحــالات، يُنتج الأفراد البيانــات ويقومون بإدخالها للعمل على الاســتفادة من خدمة (مثل وسائل التواصل الاجتماعي)، أو المشاركة في مشروع جماعي مثل مشروع خريطة الشارع المفتوح (http://www.openstreetmap.org) أو موسوعة الإنترنت ويكيبيديا (https://www.wikipedia.org). ويُســمى مثــل هــذا العمل بالإنتــاج الاجتماعي (Prosumption) حيــث تختلــط أنماط الإنتاج والاســتهلاك جزئيــاً بعضها في بعض، مع اكتســاب الأفــراد دوراً في إنتاج الخدمة أو المنتج الذي يســتهلكونه (Ritzer and Jurgenson 2010). فعلى سبيل المثال، يجري إنتاج واستهلاك محتوى مواقع التواصل الاجتماعــي في ذات الوقت من قبل الأفراد المســتخدمين عن طريق إدخال التعليقات، تحميل الصور ومقاطع الفيديو، والمشــاركة في مناقشــة وتبادل المشــاعر («الإعجاب» أو «عــدم الإعجاب» للشيء). ومن خلال جهودهم الجماعية، يُنشــئ الأفراد المنتجات والخدمات التي يشاركون فيها ويضيفون قيمة لها ويستخرجون القيمة منها، بحسب ما أشار إليه Benkler (2006) بمصطلح «ثروة الشبكات». وفي معظم الأحوال، يقوم أفراد





الإنتاج الاجتماعي بهذا العمل الإضافي بمقابل مادي قليل أو دون أي تعويض، وذلك إما للحصول على متعة المهمة، أو الشعور بالسلطة، أو توفيرهم المال والوقت مع انخفاض تكلفة الخدمة والتي غالباً ما تصبح أكثر مرونة بطبيعتها مثل الخدمات المصرفية عبر الإنترنت أو الحجز في رحلة طيران من المنزل. في المقابل، يتلقى مزوّد الخدمة أو متاجر التجزئة عملاً دون تكلفة، بالإضافة إلى الخبرات والآراء والمعرفة واكتساب الكفاءة ومعلومات ذات قيمة من خلال القرب أكثر من العملاء / الزبائن والقدرة على تسييل بياناتهم (Ritzer and Jurgenson 2010; Dodge and Kitchin 2013). فبعض أشكال الإنتاج الاجتماعي هي أقرب إلى الخدمة الذاتية ومقيدّة بطبيعتها، مع قيام أفرادها بالأعمال التي كانت مسندة إلى موظفين، في حين أن البعض الآخر أكثر انفتاحاً وتفاعلية. وهنا، سيتمّ مناقشة خمسة أشكال للبيانات التطوعية باختصار: المُعاملات (Transactions)، ووسائل التواصل الاجتماعي (Social Media)، والمراقبة الحية (Sousveillance)، والتعهيد الجماعي (Crowdsourcing)، وعلم المواطنة (Citizen Science).

المُعاملات (Transactions):

يتطلب الشراء عن طريق الإنترنت من المشتري إدخال البيانات الشخصية له فيما يخصّ عملية الدفع والعنوان وغيرها من البيانات الوصفية الضرورية الأخرى باعتبارها جزئياً عملية للتحقق وتبادل المعلومات ولكن أيضاً لمعرفة المزيد عن المستهلك. وعادة ما تكون التفاعلات في مثل هذه التبادلات محدودة إلى حدّ ما، وتقتصر على اختيار ماهية بنود الشراء، والعديد من حقول البيانات الإلزامية لوضع اللمسات الأخيرة وإتمام عملية الشراء. وبعد تبادل المعلومات، من الممكن طلب المزيد من المعلومات من المستخدمين، إما لتقييم أداء الموقع وخدمة العملاء، أو لتصنيف ومراجعة بيانات البند الذي تمّ شراؤه. وتضيف بيانات المراجعة هذه قيمة للموقع من خلال اقتراح كيفية تحسين التصميم وكذلك تزود المستهلكين الآخرين بردود الأفعال (Feedbacks) عن مدى جودة وكفاءة المنتجات. فعلى مواقع إلكترونية مثل (http://www.tripadvisor.com)، يمكن لأفراد الإنتاج المجتمعي تقييم واستعراض الفنادق وخدمات السفر الأخرى، مثل البيانات التي يجري تقديمها طوعاً على الموقع عن حركة المرور، وبثّ الإعلانات والإشارة إلى مواطن الربح المحتملة، مما يكون

لـه تأثير ملحوظ في اختيارات المسـافرين الآخرين. كما يوفّر الموقـع بيانات مفيدة حول الشـخص الذي تطوّع لإجراء الاسـتعراض والتقييم، مثل خيارات أسـلوب الحياة والإنفاق خلال السـفر التي يمكن تسـييلها مالياً عن طريق بيعها إلى أطراف ثالثة. وبالمثل، تتطلب تعبئة النماذج الإلكترونية الحكومية من المواطن تقديم معلومات للموقع بشـكل تطوعي. وهنـا، يكون هنالك تبادل للبيانـات يتعدى حدود عملية الدفـع للمبالغ النقدية، فبعض البيانات الإضافية الأخرى يمكن أن تكون مرتبطة بالمعاملات، مثل بيانات مسـار النقر على الصفحات ومعلومات بطاقة الولاء، مما يتيح بناء ملف مبدئي للشخص يمكن البناء عليه مع مرور الزمن أو يتم ربطه مع مصادر البيانات الأخرى مثل بيانات التوزيع الجغرافي للسكان.

التواصل الاجتماعي (Social Media):

قبل العام ٢٠٠٣م، كانت شـبكة الإنترنت إلى حدّ كبير وسـطاً للبثّ الإعلامي يمكن من خلالهـا للناس تقديم المعلومات أو شراء البضائع. وكانت هناك لوحات الإعلانات والقوائم البريدية التي يستطيع الناس من خلالها التواصل، ولكنها كانت قائمة على استخدام النصوص فقط، ومحددة بطريقة عرض بدلاً من كونها وسـيلة مفتوحة وعامة. وبينما يمكن للأفراد إنتاج أشـكال أخـرى من المحتوى مثل صفحات الويب، إلا أن ذلك يتطلب المهارة لإنشـاء مثل هذه الصفحات وشراء اسـم النطاق، وكانت هذه الصفحات ثابتة إلى حدّ كبير وينبغي تحريرها بشـكل رئيسي على الطريقة اليدوية ليجـري تحديثها. واعتباراً من العام ٢٠٠٤م، وعلى الرغم من ذلك، بدأت طبيعة شـبكة الإنترنت بالتغيّر لتصبح أكثر تفاعلية وتشـاركية وديناميكية، وقد تمّ الإشـارة إلى هذا التحول على أنه الانتقال من إصدار شـبكة الويب الأول (Web 1.0) إلى الإصدار الثاني (O'Reilly 2005) (Web 2.0) .

وعوضاً عن محتوى الويب الذي يجري تقديمه إلى حدّ كبير من قبل المختصين ومصممي المواقـع الإلكترونية، فقد أصبح لأي شـخص إمكانية إنشـاء محتوى يمكن تحديثه بشـكل مسـتمر. وعـلاوة على ذلك، يمكـن الوصول إلى هـذا المحتوى والتفاعـل معه من خلال مجموعـة متنوعة مـن المنصات مثل الهواتف الذكية والأجهـزة اللوحية. وبدلاً من مجرد تصفّح المعلومات أو تقييدها في مناقشـات ضيقة نسبياً، أصبح من الممكن للناس المشاركة بسـهولة وبشكل فعّال في إنتاج عالم الإنترنت وتصميم وتحرير المحتوى، والإسهام بحرية في

الخدمات التي تتبادل المواد مع الآخرين. وفي حين تم اعتبار الإصدار الأول لشبكة الويب (Web 1.0) وسطاً للقراءة فقط، فإن الإصدار الثاني لشبكة الويب (Web 2.0) هو وسط للقراءة والكتابة، يمكن للناس من خلاله إضافة قيمة للمواقع عندما يقومون باستخدامها.

وتميل مواقع الإصدار الثاني من الويب وخدماتها إلى التركيز على النشر من العديد – إلى – العديد، والتواصل الاجتماعي، وتوفير معلومات أكثر ثراء واتساقاً. وتشتمل الأمثلة على خدمات الشبكات الاجتماعية على سبيل المثال الفيسبوك وتويتر؛ والصور ومواقع مشاركة الفيديو على سبيل المثال، فليكر ويوتيوب؛ والمدونات؛ وعمليات المزج باستخدام واجهات برمجة التطبيقات المفتوحة (APIs) لدمج البيانات من مصادر مختلفة لإنشاء تطبيقات جديدة على سبيل المثال، فورسكوير، وييلب (Yelp). وتعتمد هذه المواقع كلها على المشاركة الفعالة من جمهور متحمس لتبادل المعلومات حول حياتهم والقيام بأعمال من قبيل الكتابة، التحرير، التطبيق، التعديل، النشر، المشاركة، الوسم، التواصل، ... إلخ (Beer and Burrows 2007). وعند القيام بذلك، فهم لا يتطوعون بالعمل فقط، ولكن بالكثير من البيانات عن أنفسهم، متضمناً ذلك الملفات التعريفية، الصور، المواقع، الآراء، التفضيلات، القيم، وشبكة اتصالاتهم الاجتماعية. ولأنه قد جرى تملك هذه المواقع من قبل مجموعة شركات تأسست حديثاً، فإنه يجري تداول البيانات في أيدي مملوكة للقطاع الخاص والتي تسعى بدورها لإنتاج نماذج جديدة من تراكم رأس المال عن طريق استخراج القيمة منها (انظر الفصلين الثاني والسابع).

المراقبة الذاتية (Sousveillance):

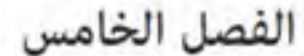
المراقبة الحية هي الرصد الذاتي وإدارة صحة الفرد الشخصية وحياته من خلال التقنيات الرقمية الحميمة على سبيل المثال، معدات اللياقة البدنية، والحوسبة القابلة للارتداء لتسجيل البيانات المتعلقة بالفرد (Mann et al. 2003). وعلى النقيض من المراقبة الإشرافية (Surveillance)، والتي يتم فيها مراقبة الفرد من موقع خارجي بواسطة شخص آخر، يجري توظيف وضبط المراقبة الذاتية عن وعي وإدراك الفرد لتحقيق الذات، وتوفير منظور داخلي أوّلي عن حياته. وعلى مدى العقد الماضي، تطورت حركة المراقبة الذاتية من أناس يراقبون ويسجّلون بنشاط بياناتهم الشخصية والتي تُعرف أيضاً باسم حركة القياس



الكمّي للذات أو التتبع الذاتي كمياً (Quantified Self). وبشكل عام، يراقب الأفراد جوانب الصحة واللياقة البدنية، ويجمعون بيانات الاستهلاك (على سبيل المثال، المواد الغذائية / السعرات الحرارية)، والحالات الجسدية (مثل ضغط الدم والنبض)، والحالات العاطفية (على سبيل المثال، المزاج ومستويات الإثارة) والأداء (على سبيل المثال، أميال المشي/ الركض / الدوران المقطوعة، وعدد ساعات النوم، وأنواع النوم)، بالتعاون مع عدد من الشركات التي تقدّم التقنيات والخدمات المرتبطة بالمراقبة الذاتية. وتشمل شركات مراقبة الأنشطة الرياضية والنوم، على سبيل المثال، شركات من أمثال فت بت، ونايك بلس، فيول باند، وجوبون (Fitbit, Nike+, Fuelband, and Jawbone) والتي تعمل على مزامنة البيانات الناتجة مع تطبيقات متابعة وتعقّب تسمح للمستخدم برصد وتحليل بياناته الشخصية. ولا يجري بالضرورة مشاركة البيانات التي تمّ إنتاجها على نطاق أوسع، على الرغم من إمكانية حدوث ذلك، ولكن يجري مشاركتها مع الشركات المزوّدة للخدمة، مما يوفّر لهم اطلاعاً نافذاً على البيانات الشخصية.

وتُعد هذه التقنيات في المرحلة الأولى من التطوير وهناك رؤى أكثر شمولاً لتسجيل بيانات الحياة (Life Logging) التي تعمل على إنشاء سجل رقمي موحّد لخبرات الأفراد، من خلال جمع متعدد الطرق عبر أجهزة استشعار رقمية، وتخزينها بشكل دائم كأرشيف شخصي متعدد الوسائط (Mann et al. 2003)، وذلك مع عدد من نماذج الأبحاث التي يجري تطويرها حالياً. وتهدف سجلات بيانات الحياة إلى إنشاء سجل متصل، وقابل للبحث والتحليل لماضي الشخص الذي يشمل كل عمل، أو كل حدث، أو كل محادثة، أو كل موقع تمّت زيارته، أو أي تعبير مادي متصّل بحياة الفرد، فضلاً عن الظروف الفزيولوجية داخل الجسم والأوضاع الخارجية المحيطة به (على سبيل المثال، الوُجهة، ودرجة الحرارة، ومستويات التلوث) (Dodge and Kitchin 2007b) - "أي مجمل المعلومات التي تتدفق خلال حياة الإنسان" (Johnson 2003: 85). ومن الواضح أن إنتاج هذه السجلات عن الحياة يثير عدداً من الأسئلة بشأن الخصوصية، وملكية البيانات المنتجة، وطريقة استخدام هذه البيانات (Dodge and Kitchin 2007b).





التعهيد الجماعي (Crowdsourcing):

التعهيد الجماعي هو الإنتاج المشترك لوسائط الإعلام، والأفكار، والبيانات من خلال العمل التطوعي من قبل العديد من الناس لحل مهمة معينة. وبينما يمكن القول بأن محتوى وسائل التواصل الاجتماعي هي نتاج تعهيد جماعي بمعنى أنه يتمّ الحصول عليه من قبل عدد كبير من الناس، إلا أن الغرض من هذه المحتوى موزّع وواسع الانتشار ويفتقر إلى التركيز. فبدلاً من ذلك، يركّز التعهيد الجماعي على الإنتاج التعاوني للمعلومات وإنشاء حلول لقضايا معينة من خلال الاعتماد على الطاقة، والمعارف، والمهارات، والعمل التوافقي والجماعي لحشد من الناس (Howe 2008). وقد بيّن (2008) Howe أن هناك أربعة تطورات تدعم نمو التعهيد الجماعي: نهضة الهواة (بحسب المعايير المهنية في الغالب)، وظهور حركة البرمجيات مفتوحة المصدر، وزيادة وفرة أدوات الإنتاج خارج إطار الشركات المصنعة، وصعود مجتمعات محلية حيوية ومنظمة وفقاً لاهتمامات الناس وميولهم على شبكة الإنترنت، فهذه التطورات مجتمعة مكّنت من تجميع موزّع وديمقراطي للأفكار والأدوات والمواد. وهناك ثلاثة أنواع رئيسية من التعهيد الجماعي التي تعمل في الوقت الحاضر: تلك التي تنتج حلاً بشكل جماعي على سبيل المثال، خريطة الشارع المفتوح (http://www.openstreetmap.org)، ويكيبيديا (https://www.wikipedia.org)؛ وتلك التي تستخدم الجمهور لتقييم حلول أو منتجات متعددة على سبيل المثال، موقع شركة أمازون (https://www.amazon.com)، ومستشار الرحلات (https://www.tripadvisor.com)؛ وتلك التي تسعى إلى حل واحد أو أكثر من الجمهور على سبيل المثال، إنّوسنتف (https://www.innocentive.com/)، والملتقيات المكثّفة لمطوري البرمجيات (Hackathons).

وبخصوص النوع الأول، أصبح إنتاج المعلومات (خريطة مفتوحة المصدر، موسوعة قابلة للتعديل والوصول إليها بحرّية) تعاونياً، وعملاً من إنتاج الأفراد، مع إعطاء المشاركين الحقوق والتسهيلات لتعديل وتنقيح وإسهاب الشرح لمساهمات الآخرين (Benkler 2006). فمن خلال العمل عبر شبكة الإنترنت يمكن لأي أحد له قدرة النفاذ إليها من أن يسهم ويضيف إلى مجمّع الخبرات المحتملة ونطاق المشاريع مقابل تكلفة أقل، وضمان ممارسة مختلف وجهات النظر والتقنيات على المشروع (Weinberger 2011). وفي الواقع،



فإن إنتاج خريطة مفصّلة للعالم أو موسوعة شاملة مع عدد قليل من المهنيين هي مهمة صعبة ومكلفة، وببساطة شاقة للغاية بالنسبة لمجموعات صغيرة من المتطوعين. ولكن بالنظر إلى آلاف المشاركين، يمكن تجزئة المهمة عليهم وحلّها. وسيكون المقابل الذي سيجنيه المشاركون هو إنتاج بيانات مفتوحة المصدر خالية من حقوق الطبع والنشر والرخص المكلفة (كما هو الحال مع بيانات الوكالة الوطنية لرسم الخرائط) (Dodge and Kitchin 2013). وبالمثل، يمكن لشركة أمازون وغيرها من شركات التجزئة على الإنترنت، من خلال تمكين المستخدمين من استعراض المنتجات التي يشترونها، استخدام قوة الجماهير لمراجعة ومشاركة آرائهم حول هذه المنتجات. كما أن الفائدة التي تعود على المشاركين هي تبادل الرأي فيما يتعلق بالمزايا النسبية للسلع والخدمات.

ويُعد استخدام المسابقات لتحفيز الجماهير على اقتراح حلول مقابل جائزة من أشكال التعهيد الجماعي الخاصة والذي تستخدمه بعض الشركات لتقديم الخبرات والأفكار (Weinberger 2011). فعوضاً عن اللجوء للمناقصات والاستعانة بمجموعة محدودة من الخبراء المحتملين، تفتح شركات مثل إنّوسنتف (Innocentive) وتوب كودر (TopCoder) المشكلة من خلال الإعلان عنها لأي شخص مهتم، وتقدّم في العادة من عشرة آلاف إلى مئة ألف دولار، وأحياناً أكثر من ذلك بكثير، لإيجاد حلول لمشاكل العملاء. وهنا، يكون الجمهور هو مصدر الحل، ولكن لا يعمل الأفراد بعضهم مع بعض بشكل جماعي، وتكون الفائدة بالنسبة للشركات هي خيارات الحل المقدمة وللمشاركين هي العائد المالي إذا تمّ اختيار الحل الخاص بأحدهم. وعلى نحو مماثل، تسعى مسابقات بيانات التطبيقات المفتوحة التي تعقدها المجالس البلدية إلى تقديم مكافآت أكثر تواضعاً مقابل تطبيقات أكثر فائدة وإثارة للاهتمام. ويمكن لهذه المسابقات أن تأخذ شكل الملتقيات المكثّفة لمطوري البرمجيات (Hackathons) التي يعمل الناس فيها معاً في فرق لتطوير حلول تطبيقات، وعادة ما تمتد على مدى يوم أو خلال عطلة نهاية الأسبوع، وتتمثل الفائدة العائدة على البلدية في استخدام بياناتها من أجل الصالح العام، وشهرة المواطنين، والمشاركة المدنية.



علم المواطنة (Citizen Science):

علم المواطنة هو شكل من أشكال التعهيد الجماعي حيث تعمل مجتمعات أو شبكات مــن المواطنين بصفة مراقبــين في بعض مجالات العلــوم (Goodchild 2007: 218). هنا، يقــوم الناس بإنتاج وإعداد ومعالجة الملاحظات العملية التجريبية، والقياســات التفصيلية للظواهر بالمجان، والتي تُعد، حاسمة ومفصلية، وذات قيمة حقيقية كنقاط بيانات للعلوم الصحيحــة. ويكــون علماء المواطنة ذوي فائدة بشــكل خاص من حيــث الرصد لأنهم قد يكونــون موزعــين جغرافيــاً ومُدرجين ضمناً في المــكان عبر الزمن حيث مــن الصعب على مســتويي التنسيق والتكلفة استخدام مساعدي البحوث المهنيين لجمع القياسات الميدانية. ومن الأمثلة على ذلك، قياســات هواة الرصد الجوي الناتجة من محطات الحدائق الخلفية للأرصاد الجوية، وهواة علم الفلك الذين يراقبون السماء ليلاً بدقة. وقد أصبح علم المواطنة راسخاً خلال القرن العشرين، وذلك دعماً لرواج العلم وتعميمه، ولكن لا يزال النشاط ضئيلاً ويعتمــد على درجة معقولة من المهارات والمعارف والمعــدات المتخصصة مثل مرقاب ذي نوعية جيدة، وفوق ذلك قدر كبير من التحفيز والالتزام المستمرين.

وقد صنّف (Bonney et al 2009; cited in Miller-Rushing et al. 2012). أشكال علم المواطنة في ثلاث فئات:

- التساهمية (Contributory): وتصمم بشكل عام من قبل العلماء ولأجلها يسهم الأفراد عامة بالبيانات في المقام الأول.
- التعاونية (Collaborative): وتصمم بشــكل عام من قبل العلماء ولأجلها يسهم الأفراد عامة بالبيانات، ولكن قد يساعدون أيضاً على تحسين تصميم المشروع، وتحليل البيانات، ونشر النتائج.
- الإنشــاء المشــترك (Co-created): وتصمم من قبل العلماء وأعضــاء من عامة الجمهور يعملــون معــاً ولأجلها يتم إشراك بعض عامة المشــاركين على الأقــل في معظم أو جميع خطوات العملية العلمية؛ ويشــمل ذلك أيضاً البحــوث التي وضع تصورها وتمّ تنفيذها من قبل علماء هواة (غير مهنيين) بشكل كامل.



وفي العادة، يختلف مســتوى الخبرة ونوع المشاركة بحسب الفئة، مع مشاركة المواطنين العلماء فقط من ذوي المهارات الأكثر تقدماً في مشــاريع الإنشاء المشترك. ويمكن أن تتخذ المشاركة عدداً من الأشكال، بما في ذلك:

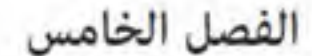
- الحوسبة الموزعة: حيث يضيف أحد المشاركين قوة معالجة حاسوبية من أجهزة الحاسب الشــخصية الخاصة بــه إلى المشروع، مما يســمح بمعالجــة البيانات باســتخدام الموارد الحاسوبية منها (التساهمية).
- النسخ / التحويل / الرقمنة: تحويل شكل من أشكال البيانات إلى آخر، على سبيل المثال، تحويل وثائق مكتوبة بخط اليد إلى نص رقمي (التساهمية).
- قياســات الرصد: ينتج العلماء البيانات، ولكن يُســتخدم علم المواطنة لإجراء القياســات باستخدام أدوات محددة يجري توريدها لهم (التساهمية).
- جمــع البيانــات: يجمع أفراد المواطنين البيانات ومن ثمّ يقومون بمشــاركتها مع العلماء (التساهمية).
- تحليل الرصد: يقوم أفراد المواطنين بإجراء بعض التحليل، بما في ذلك توفير بعض التفسيرات لما قد تعنيه النتائج، بدلاً من مجرد الاكتفاء بإنتاج أو قياس البيانات (التعاونية).
- تصميم البحث وتشــغيله: يقوم أفراد المواطنين بدور نشط في تصميم وتشغيل المشروع، فضلاً عن نشر النتائج والاستنتاجات (الإنشاء المشترك).

(Open Scientist 2013)

الخلاصة:

تطرّق هذا الفصل إلى العوامل المساعدة للبيانات الكبيرة ومصادرها. وتمّ بيان أن إنتاج البيانات الكبيرة قد تيسّر بفضل التقاء خمســة ابتــكارات تقنية من بداية الألفية الجديدة فصاعداً وهي تنامي قوة المعالجة الحاسوبية، وكثافة التشبيك الحاسوبي، والحوسبة المنتشرة والحوسبة في كل مكان، والتعرّف وتحديد الهوية من خلال زيادة قابلية الفرز والقراءة آلياً، وأخــيراً التخزين الموزّع الواســع النطاق. وقد أدّت هذه التطــورات مجتمعة إلى مجموعة





متنوعــة من النظــم الاجتماعية التقنية التي تنتج البيانات الكبــيرة. وقد قدّم هذا الفصل مسار تتبّع أولي لهذه الأنظمة، حيث قسّمها إلى ثلاث فئات واسعة النقاط من أشكال إنتاج البيانات: الموجّهة، والآلية، والتطوعية. وينبغي النظر إلى مثل هذا المســار باعتباره خطوة أولية في محاولة وضع خريطة لمشــهد يتكشّــف باستمرار. وما هو مطلوب حالياً يتمثل في مجموعة مفصّلة من الدراسات التجريبية التي تدرس بشكل أكثر عمقاً الطرق المختلفة التي يتم بها حالياً إنشــاء ومعالجة واســتخدام البيانات الكبيرة. كما تحتاج مثل هذه الدراسات إلى رسم أصول التقاء هذه التقنيات المُمكنة وفكّ مجاميع البيانات التي تمّ أنشاؤها وكيف تطبق عملياً لإنتاج أشكال معينة من البيانات الكبيرة (انظر الفصلين 1 و11). وهذا يشمل تفكيــك عناصر النظام المرتبط بها (انظر الفصل 7). وعند القيام بذلك، يمكن الحصول على أفكار ورؤى أكثر عمقاً لكل نظام اجتماعي تقني ولمختلف العمليات والضوابط والسياسات المتصلة به، والعمل المتنوع الذي يقوم به في هذا العالم.

كــما ينبغي لمجموعــة مصاحبة من الدراســات أن تختبر ما قد يحــدث عند ربط النظم الاجتماعية التقنية معاً لإنشــاء أنظمة جديدة أكبر والتي من خلالها تنتج الآثار الهامة لتضخّــم البيانات. كما أن التقنيات التي تقوم عليها البيانات الكبيرة، والبيانات نفســها من خلال علائقيتها، هي أيضاً جاهزة لمثل هذه التوســعات. في الواقع، تكون مجاميــع اجتماعية تقنيــة (Socio-Technical) معقدة، مثل سلســلة متاجر التجزئة الكبــيرة كوولمارت (Walmart)، تقــوم وعلى نحو متزايد، بربــط العديد من الأنظمة مثــل إدارة سلســلة التوريــد (Supply Chain Management)، تخطيــط مــوارد المؤسســات (Enterprise Resource Planning)، إدارة علاقات العملاء (Customer Relationship Management)، إدارة المخــزون (Store Management)، وأنظمــة الأمن (Security Systems) والتي تُنتج أشكالاً من البيانات الموجّهة، والآلية، والتطوعية لإنشاء منظمات معقدة ومتطورة موجّهة بالبيانات والتي تترتب عليها آثار متنوعة على الموردين والموظفين والمســتهلكين. وبالمثل، تحاول مراكز الســيطرة والتشــغيل للمدن الذكيــة دمج أنواع مختلفة من البيانات التي تمّ جمعهــا من مصادر مختلفة في مركز واحد حيث يتم الجمع بينها وتحليلها بشــكل آني أو لحظي (Real-Time)، مما يغيّر بشكل جذري الطريقة التي تُدار وتُحكم بها التجمعات المدنية (انظر الفصل السابع).



وقد أثار مثل هذا التوسـع وإعادة الضبط العديد من الأسـئلة الاجتماعية والسياسية والأخلاقية والمعياريـة المتعلقة بأنواع الأنظمة والأماكن التي نريد العيش داخلها، وقد بدأنا بالكاد أن نسأل ونجيب عن هذه الأسئلة (انظر الفصل العاشر).



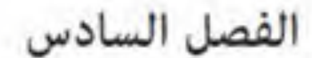


الفصل السادس

آليات تحليل البيانات (Data Analytics)

ليســت البيانات بحدّ ذاتها مفيدة، كما أنها ليســت ذات فائدة إلا إذا أمكن استخلاص معنى وقيمة منها. وبعبارة أخرى، فإن ما يمكن فعله حيال البيانات هو أمر أهم من مجرد عمليــة إنتاجها؛ إذ ترتكز جميع العلوم على تحقيق معنــى وقيمة من البيانات. إن إيجاد واستشــعار المعنى من البيانات الصغيرة الموسّــعة والبيانات الكبــيرة ينطوي على تحديات جديــدة. ففي حالــة البيانات الصغيرة الموســعة، يكمن التحدي في ربــط قواعد البيانات المتنوعــة للحصول على رؤى جديــدة، وانفتاح البيانات لتقبّل المناهــج التحليلية الجديدة التي يجري اســتخدامها مع البيانات الكبيرة. أما بالنسبة للبيانات الكبيرة، فيتمثّل التحدي في التكيّف مع الوفرة والشــمولية (متضمناً ذلك الكميات الكبيرة من البيانات ذات الفائدة والقيمــة المنخفضة)، والمناســبة الزمانية والديناميكية، والفوضى وعــدم اليقين، والعلائقية العاليــة، والطبيعة شــبه المنظمة وغير المنظمة، وحقيقة أن يتمّ إنتــاج الكثير من البيانات الكبيرة دون أي سؤال محدّد في العقل لتجيب عنه أو أنها في صحبة مُنتج هو نتيجة ثانوية لنشاط آخر. فالواقع، وحتى وقت قريب، تمّ تصميم تقنيات تحليل البيانات بشكل أساسي لاستخلاص رؤى من مجموعات بيانات شحيحة، وثابتة، ونظيفة، وضعيفة العلائقية، وجرى جمعها بطرق علمية وملتزمة بافتراضات صارمة مثل الاســتقلالية، والثبات، والعيارية، كما جرى إنتاجها وتحليلها مع وضع سؤال محدد في الاعتبار (H.J. Miller 2010).

حتى وقــت قريب كان تطوير أدوات للربط بين مجموعات البيانات المختلفة وتحليل البيانــات الكبــيرة ضعيفاً؛ بســبب التعقيــد المرتبط بتطويــر هذه الأدوات مــن الناحية الحاســوبية. ولم يجرِ تحليل مجموعات البيانات الضخمة بشكل منتظم سوى في السنوات الأربعين الماضية أو نحو ذلك. وحتى ذلك الحين كانت هذه التحليلات قائمة على مبادرات خاصة قادرة على تحمّل الموارد اللازمة. وبدون طاقة حاســوبية عالية واســعة الانتشــار ويمكن النفاذ إليها، ســيكون إيجاد واستشــعار المعنى من طوفان البيانــات الهائل مكلفاً

للغاية أو صعب التطبيق ويستغرق وقتًا طويلاً. وبالطبع، هناك نوعٌ ما من الجدل الدائر في الواقع هنا؛ إذ دون الحوسبة الواسعة الانتشار في كل مكان لم يكن إنتاج البيانات الكبيرة ليكون في المقام الأول. ومع ذلك، كما ذكر (Hastie et al. (2009: xi، فإنه بالنظر إلى «ظهور الحواسيب وعصر المعلومات، فقد تفاقمت المشاكل الإحصائية من حيث الحجم والتعقيد».

وتمثّل الحل لتحديات معالجة وتحليل البيانات الصغيرة الموسّعة والبيانات الكبيرة بمجموعة جديدة من آليات إدارة المعلومات والتخزين (انظر الفصل الخامس) وتطوير أدوات تحليل البيانات الكبيرة. فمنذ بداية الحوسبة استمرت هذه الآليات والأدوات التحليلية الجديدة في التطور، وهي تقوم على اختبارات إحصائية، ونماذج، وأساليب تمثيل صوري رصينة متّبعة، كما أنها تعمل على إنشاء أساليب وطرق جديدة تضرب جذورها في البحوث المتعلقة بالذكاء الاصطناعي والأنظمة الخبيرة التي هدفت إلى إنتاج تعليم الآلة بحيث يمكنها حاسوبيًا وبصورة تلقائية التنقيب، والكشف عن الأنماط، وبناء النماذج التنبئية. فمثل هذه الأدوات التحليلية أضحت ملائمة وبشكل مثالي في معالجة واستخلاص المعلومات من مجموعات بيانات ضخمة ومتصلة ببعضها، فقد أصبحت تشغل حيزًا كبيرًا من الاستثمار في البحوث من أجل توسيع وإنتاج طرق معالجة جديدة للبيانات والخوارزميات الإحصائية والنمذجة وتقنيات التمثيل الصورية (المؤسسة الوطنية للعلوم National Science Foundation 2012). وتسعى هذه الأدوات التحليلية، كما هي مطبّقة في مجال الأعمال التجارية والعلوم، للإجابة عن أربع مجموعات أساسية من الأسئلة (Minelli et al. 2013):

- الوصف (Description): ماذا ومتى حدث ذلك الشيء؟ وكم مرة يحدث؟
- التفسير (Explanation): لماذا حدث ذلك؟ وما هو تأثيره؟
- التنبؤ (Prediction): ماذا يمكن أن يحدث بعد ذلك؟ وماذا لو فعلنا هذا أو ذاك؟
- العلاج (Prescription): ما الحل أو النتيجة المُثلَى؟ وكيف يمكن تحقيق ذلك؟



وتُستمد الإجابة عن هذه الأسئلة من أربع فئات رئيسة من آليات التحليل: التنقيب عن البيانات والتعرّف على الأنماط، التمثيل الصوري للبيانات وآليات التحليل البصرية، التحليل الإحصائي، والتنبؤ والمحاكاة والتحسين. وسيتم مناقشة كل من هذه الفئات باختصار، ولكن ينبغي تقديم مرحلة ما قبل التحليل وتعلّم الآلة أولاً، لأنها مركزية لجميع الفئات الأربعة.

ما يسبق التحليلات (Pre-analytics):

تحتاج كل تحليلات البيانات إلى أن تكون البيانات المطلوب تحليلها معدّة سلفًا للتحليل؛ بمعنى أنه قد جرى تجهيزها والتحقق منها، وقد بيّن H.J. Miller (2010) and Han et al. (2011) أربع عمليات فيما يتعلّق بالبيانات الموسّعة والكبيرة والتي تطبّق عادة بالتسلسل، على الرغم من أنه لا يلزم تنفيذها بأي ترتيب معين، إضافة إلى إمكانية تكرار تنفيذها:

- اختيار البيانات (Data selection): تحديد مجموعة جزئية من المتغيرات الأكثر قيمة، وإطار العينة المحتمل لهذه المتغيرات. حيث لن تكون كل المتغيرات الناتجة ذات أهمية، ولا بالضرورة كل نقاط البيانات، فإدراج مثل هذه البيانات غير الضرورية في نماذج قد يولّد التكرار وعدم الكفاءة.
- المعالجة المُسبقة للبيانات (Data pre-processing): تنظيف البيانات المختارة لإزالة الضوضاء أو الأخطاء أو التحيّز، أو التعامل مع الحقول المفقودة أو التناقضات وعدم السلامة، وهيكلة البيانات لإدخالها في عملية التحليل.
- اختزال البيانات وإسقاطها (Data reduction and projection): التقليل من أبعاد البيانات من خلال التحويلات (على سبيل المثال، التسوية والصقل (Smoothing)، وبناء السمات والخصائص، والتجميع، والتطبيع، والتسلسل الهرمي للمفهوم، والأساليب الإحصائية مثل تحليل مستوى الانحدار وتحليل المكونات الرئيسية) لمكافأة تمثيل البيانات ولكن بصورة أكثر كفاءة.
- إثراء البيانات (Data enrichment): دمج البيانات المختارة مع بيانات أخرى (مثل بيانات التعداد السكاني وبيانات السوق) من أجل الوصول إلى رؤى يمكن الاستفادة منها بشكل أكبر.

فمــن جهة، صُمّمــت كل خطوة من هذه الخطــوات من أجل زيــادة نوعية البيانات المســتخدمة في التحليل، ومن جهة أخرى، باســتثناء إثراء البيانات، مــن أجل تقليل حجم البيانــات التي عــلى التحليل التعامل معهــا. فالأول تمّ تصميمه لتعزيــز صحة وصلاحية الاســتنتاجات المستخلصة بالنظر إلى الفوضى الكامنة وعدم اليقين في البيانات الكبيرة (انظر الفصل التاســع)، في حين يهدف الثاني إلى زيادة الكفاءة التشــغيلية مع الأخذ بعين الاعتبار أحجام البيانات وسرعاتها والأحمال الحاســوبية. ومن الممكن أن يكون من الصعب القيام بهذه المهام عند إجراء التحليلات بشكل آني أو لحظي، وخاصة فيما يتعلّق بتنظيف البيانات. ونتيجــة لذلك، وفيما عدا التمثيل الصوري للمؤشرات، لا يجري تحليل البيانات الكبيرة آنياً أو بشــكل لحظي، ولكن يجري التحليل على سلاسل تمثل عينات كبيرة جدًا لبيانات نظيفة، ومختزلة، ومخصّبة (ثرية) وخلال أزمان مختلفة. أما إذا كان المطلوب تحليلات بشكل آني أو لحظي، فتطبّق في العادة التحليلات المســبقة للبيانات مقدّمًا على عينة من أجل تحديد طبيعة البيانات وكيفية التعامل معها للانتقاء، والاختزال، والتنظيف على عجل.

ويعــد إثراء البيانات مهمة ذات قيمة عالية لأنها تنتج أثرًا مضخّمًا (Crampton et al. 2012) يُمكّــن من اكتســاب رؤى لا يمكن الحصول عليها من خــلال قاعدة بيانات واحدة. وبحســب من تمّت مناقشــته في الفصــل الثاني، فإن المــبرر الرئيسي لإنشــاء البنى التحتية للبيانات تمثل في إعادة استخدام البيانات. وإن التحدي يتمثل في إنشاء طرق لربط البيانات التــي أنتجت لأغراض مختلفة بعضها ببعض، والتي قد تختلــف أيضًا في البيانات الوصفية المصاحبة لها، ومعايير البيانات، ووحدات القياس، والإجراءات، والفئات، والنطاق، والتزامن، وتنســيقات الملفات، أو إنشــاء طرق للربط لتقليل حدوث المغالطات المحتملة الناتجة عن الاستدلال (Creation of Potential Ecological Fallacies). وهذه ليست مهمة يسيرة، ولكن تعمل الأساليب الحاســوبية على جعلها أسهل من خلال استخدام الخوارزميات التي يمكنها البحث، المطابقة، الدمج، إعادة الجمع من خلال أنواع مختلفة من التحولات، وإعادة تهيئة البيانات. وبالنتيجة، يمكن التنقيب والتحليل في مجموعة بيانات جديدة باســتخدام الفئات الأربع الرئيسية للتحليلات التي سيجري نقاشها أدناه.



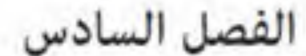
ويمكن أن تكون أعمال التحليلات المسبقة مملّة باستغراقها وقتًا طويلاً للغاية، ولكنها مع ذلك مهمة جدًا ولا يمكن تجاهلها. وبالنظر إلى الزيادة المنقطعة النظير في مختلف أنواع البنى التحتية الجديدة للبيانات والبيانات الكبيرة، فقد أصبحت منطقة ثرية للبحث حيث يسعى علماء البيانات إلى إيجاد طرق أكثر إنتاجية وكفاءة وفعالية لإتمام، وخصوصًا أتمتة، هذا العمل.

تعليم الآلة أو التعلّم الآلي (Machine learning):

يمكن إجراء تحليل الأعداد الكبيرة جدًا من سجلات البيانات في إطار زمني مناسب عن طريق الخوارزميات الحاسوبية. وفي حين أن الكثير من عمليات تحليل البيانات الكبيرة يمكن إجراؤها بالطرق نفسها المتبعة في تحليل البيانات الصغيرة، حيث يتخذ المحللون القرارات المتعلقة بالكيفية التي يتم فيها الاستكشاف من خلال هذه الخوارزميات، فإن الهدف من أغلب البحوث هو تطوير عمليات مؤتمتة يمكنها تقييم البيانات والتعلم منها ومن تحليلها. ويطلق على هذه العمليات المؤتمتة مصطلح تعلّم الآلة أو التعلّم الآلي وتشكّل أحد فروع الذكاء الاصطناعي. ويهدف تعلّم الآلة إلى التطوير المتكرر لفهم مجموعة البيانات؛ والتعلم التلقائي لإدراك أنماط معقّدة وبناء نماذج توضح وتتوقّع مثل هذه الأنماط؛ وتحسين النتائج (Han et al. 2011).

ويتكوّن تعلّم الآلة بشكل عام من نوعين رئيسيين: مُراقب (باستخدام بيانات التدريب) وغير مراقب (باستخدام التنظيم الذاتي). في التعلّم المُراقب، يتمّ تدريب نموذج لمطابقة المدخلات مع مخرجات معلومة. على سبيل المثال، يمكن تدريب النموذج لمطابقة الرموز البريدية المكتوبة بخط اليد مع مثيلاتها المطبوعة آليًا، أو التنبؤ بنتائج معينة، فهي «مراقبة» بمعنى أن بيانات التدريب موجودة لتوجيه عملية التعلّم (Hastie et al. 2009). في المقابل، تسعى نماذج التعلّم غير الخاضعة للرقابة إلى تعليم نفسها على اكتشاف الأنماط وإيجاد هياكل البيانات دون استخدام بيانات التدريب. وبشكل عام، يمكن تحقيق ذلك من خلال تحديد مجاميع وعلاقات بين البيانات التي لم يكن معروفًا بينها مسبقًا خصائص للتشابه أو الارتباط. فعلى سبيل المثال، يُمكن أن يتعلم النموذج على كيفية تقسيم العملاء إلى مجموعات ذاتية متماثلة والتنبؤ بالمشتريات لتلك المجموعات (Han et al. 2011). في كلتا الحالتين، يتم





إنشـاء نموذج من خلال عملية التعلم التي شـكلتها القواعد والأوزان التي توجه كيفية بناء نموذج فيما يتعلق بالبيانات (Hastie et al. 2009). وتبدأ عملية بناء نموذج من بناء بسيط ثم يتم التعديل مرارًا وتكرارًا على ذلك البناء باسـتخدام قواعد التعلم، بالمثل كما يُطبق مع الطفرات الجينية، إلى أن يتطور البناء إلى نموذج قوي (Siegel 2013: 122). وهناك شـكلان للتعلم المُراقب وغير المُراقب وهما التعلم شـبه المُراقب الذي يشـتمل على استخدام كلٍّ من التدريب والبيانات التي لا تحمل عناوين (وهي البيانات التي يُمكن الحصول عليها بسـهولة ولكنها غير مسـماة كالصور والتسـجيلات الصوتية، والفيديو، المقالات الإخبارية، وغيرها من البيانات الخالية من أي ملصقات تتيح التعرف عليها بشكل أكبر)، والتعلم النشط الذي يتيح للمستخدمين القيام بدور نشط في توجيه نموذج التعلم (Han et al. 2011).

وقـد جرى اسـتخدام التعلّم الآلي في جميع عمليات تحليـلات البيانات الكبيرة الأربع، عـلى الرغم من أن هذه التحليلات ليسـت مقتصرة في عملها على هـذه الطريقة. ويتمّ إجراؤهـا، في كثير من الحالات، من قبل الباحث بشـكل مباشر، وحتى ضمن التعلم الآلي، يبقى المحلّل ودوره مهمًا في الحكم على العملية وتوجيهها وتقييم النتائج الوسيطة. وكما لاحـظ H.J. Miller (2010) فـإن التعلّم الآلي ليس مجرد علم تلقائي وسـهل المنال، بل يتطلب خبرة بالمجال وفكرًا عميقاً؛ وهي المهارات التي لا يزال العقل البشري أفضل بها من أجهزة الحاسـب الآلي. وبينما تم إحراز تقدّم كبـير في تطوير تقنيات التعلّم الآلي، فإنه لا يزال من العلوم الناشـئة التي مازالت بحاجة إلى الكثير من البحث لتحسين فعالية ومتانة النماذج المنتجة.

وفيما يلي، سـيتمّ مناقشـة كل فئة من فئات التحليلات الرئيسية على حدة، على الرغم من أنه تجدر الإشارة إلى أنها غالبًا ما تستخدم جنبًا إلى جنب. على سبيل المثال، ينبغي أن يوفّر التنقيب عن البيانات والتعرف على الأنماط أسـس التنبؤ/التوقع أو التحسـين، ويُمكن اسـتخدام الإحصاءات في التنقيب عن البيانات للكشـف عن الأنماط أو في حساب التوقع/التنبؤ، وقد يسـتخدم التمثيل الصوري في التنقيب عن البيانات، أو لتحصيل المخرجات من أنظمة المحاكاة... إلخ.



التنقيب عن البيانات والتعرّف على الأنماط (Data Mining and Pattern Recognition):

التنقيب عن البيانات هي عملية استخلاص البيانات وأنماط من مجموعات البيانات الكبيرة (Manyika et al. 2011). ويقوم التنقيب عن البيانات على مفهوم امتلاك كل مجموعات البيانات الضخمة معلومات مفيدة ذات معنى على صورة غير عشوائية، وذات صلاحية سارية المفعول، وقابلة للرواية، ومفيدة، ومفهومة في نهاية المطاف (Han et al. 2011). وبالمثل، يستخدم التعلّم الآلي المراقب وغير المراقب للكشف والتصنيف والتقسيم عن العلاقات والارتباطات والاتجاهات ذات المغزى بين المتغيرات. وهي تقوم بذلك باستخدام سلسلة من التقنيات بما في ذلك معالجة اللغات الطبيعية، والشبكات العصبية، وأشجار اتخاذ القرار، والأساليب الإحصائية (دون معاملات وذات المعاملات). وتختلف طريقة الاختيار بين نوع البيانات (المنظمة، وغير المنظمة، وشبه المنظمة) والهدف من هذا التحليل (انظر الجدول ١ - ٦)).

الجدول رقم (١ - ٦)

مهام وأساليب التنقيب عن البيانات

مهمة التنقيب عن البيانات	الوصف	الآليات
التجزئة أو التجميع (Segmentation or Clustring)	تحديد قائمة من المجموعات الضمنية التي تصف البيانات	التحليل العنقودي (Cluster Analysis)
التصنيف (Classification)	التنبؤ باسم الفئة التي تنتمي إليها مجموعة من البيانات استناداً إلى بعض مجموعات بيانات التدريب.	تصنيف النظرية الافتراضية (Bayesian Classification) استقراء شجرة القرارات (Decision Tree Induction) الشبكات العصبية الذكية (Artificial Neural Networks) آلة دعم التمييز (Support Vector Machine)

الربط (Association)	إيجاد روابط بين كائنات البيانات؛ التنبؤ بقيمة بعض الحقول بالاعتماد على قيمة حقول أخرى	قوانين الارتباط (Association Rules) تصنيف النظرية الافتراضية (Bayesian Classification)
الانحراف (Deviation)	إيجاد عناصر البيانات المعرضة للانحراف عن المتوقع	التحليل العنقودي (Cluster Analysis) كشف الانحراف (Outlier Detection) تحليل التطور (Evolution Analysis)
الاتجاهات (Trends)	الاتجاهات والمنحنيات التي تلخص قواعد البيانات، عادة عبر الزمن	الانحدار (Regression) استخلاص تسلسل النمط (Sequence Pattern Extraction)
التعميم (Generalization)	دمج مواصفات البيانات	القواعد والقوانين الموجزة (Summary Rules) الاستقراء الموجه بالسمات (Attribute-Oriented Induction)

المصدر: Miller and Han (2009: 7)

إن معظم الآليـات المدرجة في الجدول (1-6) تتصل بالبيانات المنظّمة كتلك الموجودة في قواعد البيانات العلائقية. على سبيل المثال، من الممكن تطبيق نماذج التجزئة على قواعد بيانات عملاء التجزئة ومشـترياتهم لتقسـيمهم بحسـب معرّفات مختلفة بالاعتماد على خصائصهم وأنماط سـلوكهم من أجل تقديم عروض/ خدمات مختلفة لكل مجموعة مختلفة. وفي تحليل الشبكات الاجتماعية، يُمكن تحليل الروابط بين الأفراد لفهم المتغيرات الاجتماعية وكيفية تدفق المعلومات فيما بينهم. وفي كشـف الارتباطات، يمكن اسـتخدام مجموعة متنوعة من نماذج الانحدار في الكشف عن الروابط لحساب الاقترانات بين المتغيرات، ومن ثم الكشف عن الأنماط

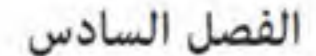


الخفية التي يمكن الاستفادة منها لتحقيق مكاسب تجارية (على سبيل المثال، تحديد ما يتمّ شراؤه كوحدة من البضائع وإعادة تنظيم المخزون لتعزيز الشرائية) (انظر الفصل السابع).

وتُثير البيانات غير المنظمة، والتي تكون على شكل نصوص أو صور أو أصوات، تحديات خاصة فيما يتعلق بالتنقيب عن البيانات. فآليات معالجة اللغات الطبيعية تهدف إلى تحليل اللغة الإنسانية التي يجرى التعبير عنها من خلال الكلمة المكتوبة والمنطوقة. وهي تستخدم دلالات وتصنيفات من أجل التعرّف على الأنماط واستخراج المعلومات من الوثائق. وتشتمل الأمثلة على استخراج العنصر الذي يستخرج البيانات الوصفية من النص بصورة تلقائية من خلال البحث عن أنواع معينة من النصوص والتعابير، مثل أسماء الأشخاص والمواقع والتواريخ والمصطلحات التخصصية والتراكيب المنطقية؛ واستخراج العلاقات التي تحدّد تلقائيًا العلاقات بين العناصر الدلالية، وربطها معًا (على سبيل المثال، اسم الشخص وتاريخ ميلاده أو مكانه، أو رأيه ببند ما) (McCreary 2009). ومن التطبيقات النموذجية لمثل هذه الآليات التحليل الوجداني الذي يسعى إلى تحديد الطبيعة العامة للآراء وقوتها حول قضية ما، على سبيل المثال، ما يقوله الناس عن المنتج على وسائل التواصل الاجتماعي. كما يمكن تعقّب أماكن نشوء هذه الآراء باستخدام البيانات الوصفية الموضعية (Graham et al. 2013) والتنقيب عن نشر المعلومات داخل وسائل التواصل الاجتماعي، على سبيل المثال، تبادل عناوين المواقع الإلكترونية الأكثر تفضيلاً ومشاركتها بين مختلف المستخدمين (Ohlhorst 2013). فهذه المعلومات مفيدة للشركات، مثل الوكالات الإعلانية، والمسوّقين، والخدمات المالية، والتي تسعى إلى الاستفادة من الاتجاهات الناشئة والوصول المنظّم في الوقت المناسب (على سبيل المثال، لوضع إعلانات تتناسب مع الأحداث الجارية؛ لبيع وشراء الأسهم قبل ردة الفعل العنيفة للسوق).

أما الصور فيتم إنشاؤها للتخزين والعرض، وليس للمحتوى والبحث (Ohlhorst 2013). فكشف وتصنيف واستخراج أنماط داخلها، مثل التعرف على الوجه أو مكان، ليس بالأمر السهل، ولكن يتم التعامل مع ذلك من خلال تقنيات المسح التصويري، والاستشعار عن بعد، ومعالجة الصور وتقنيات الاستبصار الآلية، متضمناً ذلك التعرف على الأشياء ومطابقة القالب باستخدام مجموعات التدريب، وطرق التجميع، والشبكات العصبية. وتتفاقم





مشــكلة التنقيب في الصور عند محاولة استخراج، ومقارنة، وفهرسة الأنماط لعدد هائل من الصور (Zhang et al. 2001). وبرغم أن التنقيب في الصور لا يزال من المواضيع الناشئة، إلا أنــه أصبح أكثر تقدمًا في الســنوات الأخيرة. فعلى ســبيل المثال، فــإن الموقع الإلكتروني على شــبكة الإنترنت: (http://imagevision.com) يدعي بأنه سيكون قادراً على تصنيف 50400 دقيقة من الفيديو في الســاعة الواحدة، لكل خادم، وذلك باســتخدام خوارزميات التعلم الآلي للكشف عن بعض الخصائص مثل العري وشعارات الشركات.

التمثيل الصــوري للبيانات وآليــات التحليل الصوريــة (Data Visualisation and Visual Analytics):

شاعت مقولة تنص على أن صورة واحدة تغني عن ألف كلمة. وبالمثل، جرى استخدام الســجلات البصرية لاختزال ووصف مجموعات البيانات من خلال المخططات والرسومات الإحصائية، والرسوم البيانية، والخرائط، والرسوم المتحركة. وتكشف هذه الأساليب البصرية وتوصل على نحو فعال هيكل المتغيرات، وأنماطها، واتجاهاتها، والروابط بينها. وبالنظر إلى الحجــم الهائل للبيانات الكبيرة وسرعتها، فليس من المســتغرب إذاً أن التمثيل الصوري قد صار وسيلة شعبية لاستشعار البيانات وإيصال هذا الشعور.

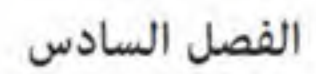
ويمكن اســتخدام التماثيل الصورية الناشئة ضمن عالم التقنية الرقمية للتنقل والاستعلام عن البيانات، مما يتيح للمســتخدمين الحصول على لمحة عامة عن كامل مجموعة البيانات، بالتركيز على البنود المهمة، واســتبعاد البيانات غــير المهمة، واختيار عنصر بياني أو مجموعة مــن البيانات والحصول على مزيد من التفاصيل، وعرض العلاقات بين العناصر، واســتخراج مجموعات فرعية من التفاصيل عند الحاجة (Shneiderman 1996). وعند عمل ذلك، يمكن الكشــف عن خصائص وبنية مجموعات البيانات، وأكثر من ذلك يمكن اســتخدامها لتوضيح وفهم النــماذج ومجموعات البيانات المعقدة التي يتعذّر تمثيلهــا بطريقة مجردة وواضحة ومفهومــة (مثل الظواهر الذرية، والظواهر الفلكية الثلاثية الأبعاد)، ورســم عشرات الآلاف من نقاط البيانات للكشف عن البنية، والتجميعات، والثغرات، والقيم الشاذة التي قد تظل بدلاً من ذلك مجهولة في الخفاء (Shneiderman 1996). على سبيل المثال، استيعاب وفهم ملايين التغريدات على شــبكة التواصل الاجتماعي تويتر ليســت بمهمة سهلة، فيمكن للمرء



الحصــول على فكرة تقريبية عن القصص أو المواضيع التي تبدو واعدة، ولكن الحصول على نظرة عامة وشــاملة هي مهمة غاية في الصعوبة، وكيفية تغيّر التوجه على مستوى الجماهير والأماكن. فأحد الحلول يتمثل في رســم خريطة تغريدات موزعة جغرافيًا ومصنّفة بحسب الميــول والأهواء، وهو نهج جرى الأخذ به من خــلال فريق عمل الموقع الإلكتروني (http://floatingsheep.org). لقد أنشــؤوا عشرات الخرائط حول توجهات المواضيع والتغريدات، واشــتمل ذلك على الشــكل (١ - ٦) والذي يعرض التوزيع الجغرافي لتغريدات التخوف من المثليــين في الولايــات المتحدة من يونيــو ٢٠١٢ م وحتى إبريــل ٢٠١٣ م. كما أنهم عرضوا محتوى موسوعة الويكيبيديا وجوجل بالنسبة لمواضع التوزيع الجغرافي حول العالم.

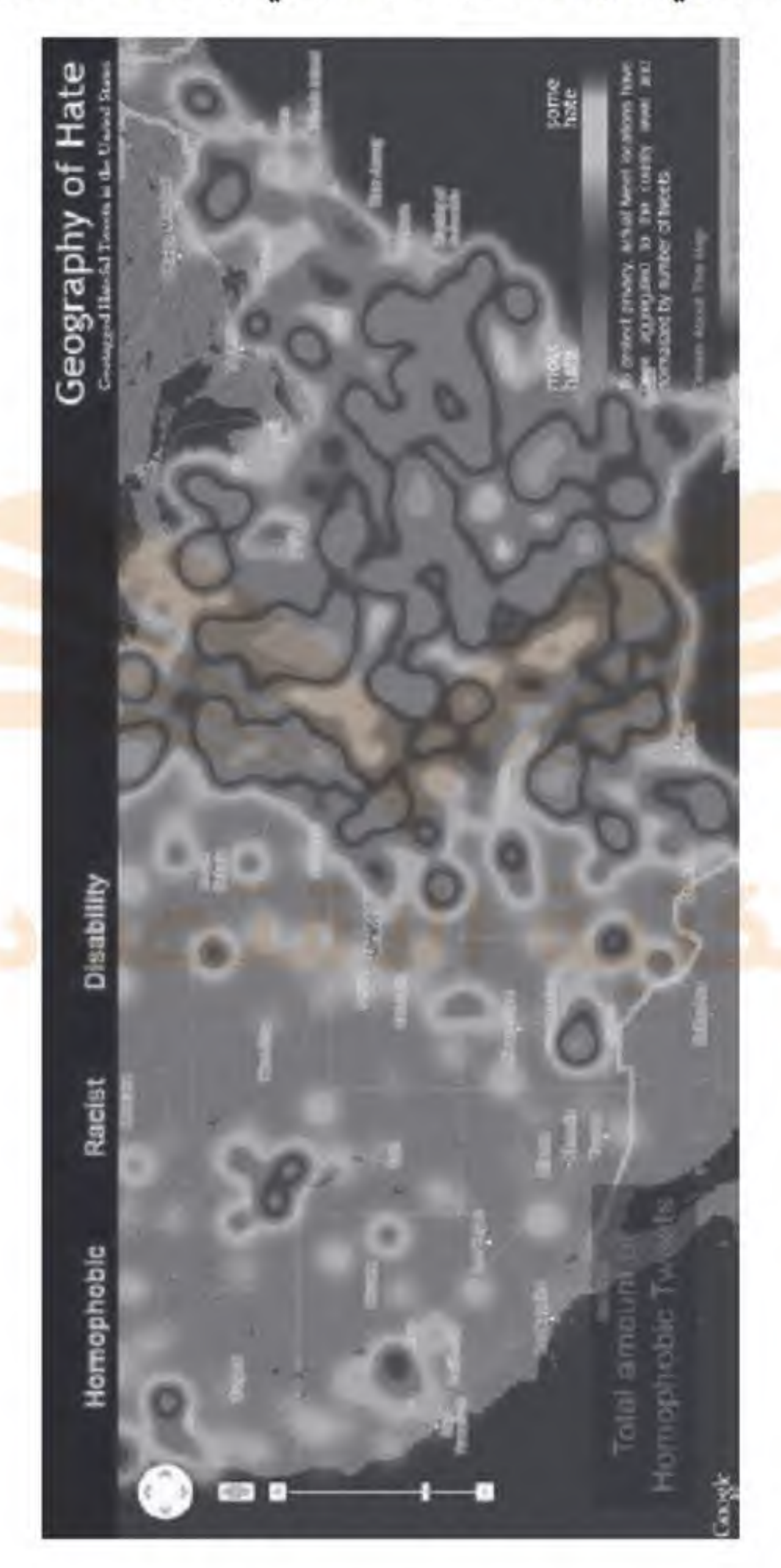
ويشيع اســتخدام التمثيل الصوري أيضًا ليكون وسيلة للمراقبة الآنية المستمرة لظاهرة ما بشــكل آني أو لحظي، مما يتيح تعقّب ورصــد العديد من المتغيرات عبر الزمان والمكان، والمقارنــة بينها، وتحديــد التغيير الحاصل. فلوحــات التحكم والعــرض (Dashboards) للبيانات المتغيرة تكون غالباً معروضة على شاشات الحاسب الآلي في غرف التحكم الحديثة، وهي تُلخّص على نحو صوري الأنظمة المتغيرة باســتمرار للمشــغلين الآدميين، مع سلاســل زمنية ورســوم بيانية وخرائط للكشــف عن الأحداث (انظر ليك Lake 2013، للمقارنة بين ٢٤ لوحة عرض). وعلى سبيل المثال، يمكن أن توفّر البيانات من مختلف أنحاء نظام النقل خريطــة لتدفق حركة المرور وتقارير الحوادث بشــكل آني أو لحظــي؛ أو مواقع الطائرات خلال مرورها عبر منطقة ما (انظر الشكل ٢ - ٦)، كما قد توفّر بيانات رادار الطقس خريطة لهطول الأمطار بشكل آني أو لحظي ومحاكاة للساعات القليلة الماضية. فمثل هذه البيانات الصورية لا تساعد ضباط إدارة حركة المرور والراصدين الجويين فقط، ولكن أيضًا المواطنين الذين يمكنهم الوصول إليها ومتابعة الأوضاع الجارية باســتخدام الحاسب الآلي أو الهواتف الذكية وتبديل ســلوكهم لتجنب طرق معينة أو التعامل بشــكل مناســب مع المستجدات. ومثــال على نموذج أوليّ للوحة تحكّم خاصة بمدينة عامة تضمّ مجموعة من البيانات الآنية أو اللحظية عن الطقس، وتلوّث الهواء، ومعوّقات حركة مرور النقل العام، وتوافر الدراجات العامة، ومســتوى ارتفاع المنســوب المائي النهري، وارتفاع الطلب عــلى الطاقة الكهربائية، وســوق الأوراق المالية، واتجاهات التغريدات على شبكة التواصل الاجتماعي تويتر، وتغذية كاميرات مراقبة حركة المرور تلك المعروضة في الشكل (٣-٦).





الشكل رقم (١ - ٦)

التوزيع الجغرافي لتغريدات التخوّف من المثليّة في الولايات المتحدة الأمريكية



المصدر: #http://users.humboldt.edu/mstephens/hate/hate_map.html



الشكل رقم (٢ - ٦)

المواقع اللحظية لمسار الرحلات الجوية



المصدر: /http://www.flightradar24.com

الشكل رقم (٣ - ٦)

لوحة معلومات مدينة لندن الخاصة بالمركز البريطاني للتحليل العمراني المتقدم

London 51.51 N, 0.13 W

Wed 14 Oct @ 08:48:17

Go to Map - Go to Grid - Change City

WEATHER STATION (CASA TEAM)

STATION	WIND SPEED	WIND GUSTS	DIRECTION	TEMPERATURE	HUMIDITY	RAIN TODAY	PRESSURE	FORECAST
CASA Office Bloomsbury W1	2.0 mph	13.0 mph	W →	8.2 °C	81%	0.0 mm	1022.55 mbar	Cloudy
SJG Weather Pinner HA5	0.0 mph	3.6 mph	NE ↙	17.8 °C	63%	0.0 mm	1016.8 mbar	Sunny

WEATHER (METAR)

London City Airport

Mostly clear — NW at 5 mph — 7 C

FORECAST (YAHOO! WTH)

Wed 14 C Partly Cloudy — Thu 13 C Mostly Cloudy

TUBE LINE STATUS (TFL)

Line	Status
Bakerloo	Good Service
Central	Good Service
Circle	Good Service
District	Good Service
H & C	Good Service
Jubilee	Good Service
Metropolitan	Good Service
Northern	Good Service
Piccadilly	Good Service
Victoria	Good Service
W & C	Good Service
Overground	Good Service
	Good Service
DLR	Good Service

LONDON CYCLE HIRE (TFL)

2.3 % Stations Full — 5.6 % Stations Empty

9387 Bikes Available — 307 Bikes or Docks Faulty

Available Bikes (last 24h): 9500, 9000, 8500; 24h 20h 16h 12h 8h 4h now

IN SERVICE (TFL)

4245 London buses

325 Underground trains

AIR POLLUTION (DEFRA)

μg/m³ TIME AVGD	OZONE	NO_2	SO_2	$PM_{2.5}$	PM_{10}
Bloomsbury	17	39	5	10	22
Marylebone Rd	34	30	3	11	18
N Kensington	37	30	2	9	18

BICYCLES (LBH)

Goldsmiths' Row 4725 yesterday

STOCKS (YAHOO)

FTSE 100 Index 6342.28 -28.90 (-0.46%)

TRAFFIC CAMERAS (TWO AT RANDOM) (TFL)

Station Rd/Wellesley Rd — Kew Bridge Northside

BBC LONDON NEWS (BBC)

OPENSTREETMAP UPDATES (OSM)

Adding Motorcycle parkings Adding Motorcycle parkings Adding Motorcycle parkings Adding Motorcycle parkings Adding Motorcycle parkings Adding Motorcycle parkings Adding Motorcycle parkings

ELECTRICITY (N GRID)

Data unavailable

MOOD (LSE MAPPINESS)

1% happier than the long term average for here — 5% happier than the whole country right now

TWITTER TRENDS FOR LONDON

#DemDebate Lamar Odom Willian #Love2PartyNotts #Innotribe #BETHipHopAwards #EuropeSwagger

LONDON NEWS AND EVENTS (TWITTER)

Londonist: Ghost walks, horse-drawn carriage rides, and other things to do for #Halloween in #London http://t.co/R5hXmo4mAY http://t.co/F3minnWXUB

LONDON UNIVERSITIES (TWITTER)

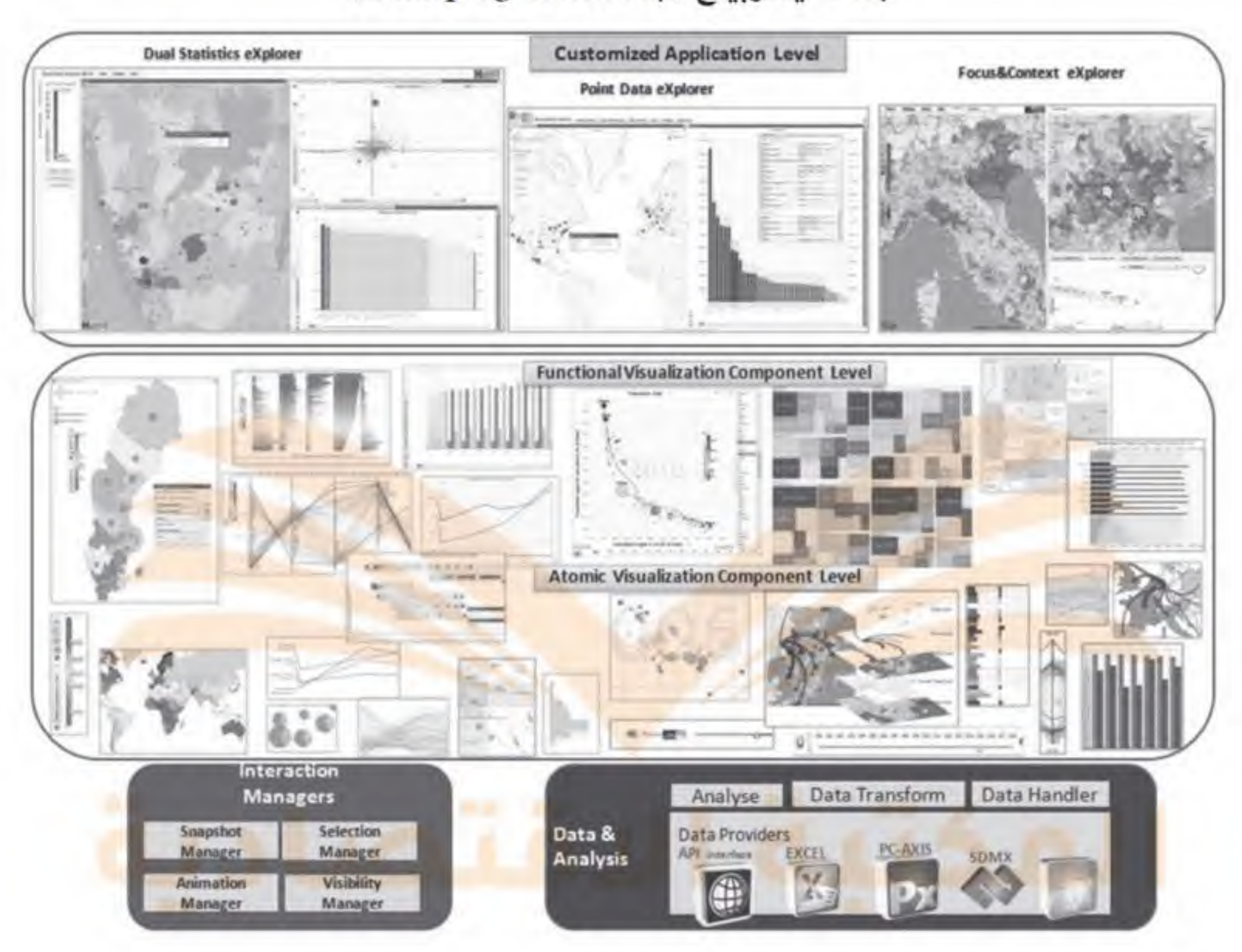
Imperial College: Tomorrow we're launching this year's #altheaimperial prize for female student entrepreneurs http://t.co/yTlzcjYEhE http://t.co/xG4RMZ3onc

المصدر: http://citydashboard.org/london/



الشكل رقم (٤ - ٦)

مجموعة أدوات التحليلات الجغرافية الصورية التي وضعها المركز الوطني للتحليلات الصورية في جامعة ينكوبينج (Linköping University)



المصدر http://ncva.itn.liu.se/tools?l=en

الشكل رقم (٥ - ٦)

استخدام مجموعة أدوات التحليلات الجغرافية الصورية في بناء الرواية القصصية الجماعية

المصدر http://ncva.itn.liu.se/tools?l=en

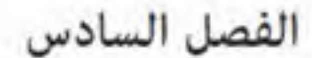


ومن الممكن أيضًا استخدام التمثيل الصوري كشكل من أشكال الاستنتاج التحليلي. وهنا، لا يصف أو يعرض التمثيل الصوري مجرد البيانات فقط، ولكن يجري استخدامها أداة تحليلية. ويتم توجيه هذه الطريقة، والتي يطلق عليها مصطلح التحليلات الصورية، في الغالب بواسطة مجموعة من الأشخاص والخوارزميات التي تعمل باتساق لاستخراج البيانات، وبناء النماذج البصرية وتفسيرها، والمضي قدمًا في المزيد من التحليل الإحصائي (Keim et al. 2010). لقد أصبحت آليات التحليل البصرية مجالاً مهمًا وجديدًا من مجالات البحث، وهو الأمر الذي أسهم في تطوير حقل التمثيل الصوري للمعلومات من خلال دمج عناصر التنقيب عن البيانات، وعلوم الإحصاء والإدراك (Thomas and Cook 2006). فالعديد من أنواع الرسومات البصرية المختلفة شائع الاستخدام، وهي في العموم تفاعلية وتتيح للمستخدم التعامل معها والتلاعب في الصور من أجل البحث والكشف عن الأنماط والروابط. وبالإضافة إلى ذلك، فإن الشاشات المعروضة للمستخدم متصلة ببعضها في الغالب بحيث إن التفاعل مع إحداها ينعكس على الأخريات، مما يمكّن من تفحّص جوانب من البيانات من أكثر من منظور واحد في ذات الوقت. فعلى سبيل المثال، يعرض الشكل (٤-٦) أنواع آليات التحليل الصورية المتاحة ضمن أدوات آليات التحليل الصورية (GAV)، الموضوعة من قبل المركز الوطني للتحليلات الجغرافية الصورية في جامعة ينكوبينج (Linköping University) في السويد. بالإضافة إلى توفيرها مجموعة من الأدوات التحليلية، تمكّن أدوات المجموعة من تحليل البيانات واستكشافها بشكل تعاوني، وبناء التطبيقات، والتداخل مع تطبيقات الخرائط كخرائط جوجل، وبناء وتبادل القصص السردية. وينشئ الأخير بُعداً اجتماعيا للأدوات، مما يمكّن الزملاء وغيرهم من تبادل تفسيرات حسية متعلّقة بالتمثيل الصوري (انظر الشكل ٥ - ٦).

التحليل الإحصائي (Statistical analysis):

هناك تاريخ طويل من الأساليب الإحصائية التي يجري تطبيقها على البيانات الكمية في محاولة لفهمها. فالإحصاء الوصفي يفصّل خصائص وتوزيع نقاط البيانات ومستوياتها من الخطأ وعدم اليقين. ويشتمل الإحصاء الوصفي على تحليل السلاسل الزمنية التي تنص على كيفية تغيّر البيانات بمرور الزمن، ونظرية الرسم البياني التي تدرس بشكل رياضيّ تنظيم





هياكل الشبكة، والإحصاء المكاني الذي يصف هندسة وأنماط التجميع، والتشتت، والانتشار، في حين يسـعى الإحصاء الاسـتدلالي إلى شرح، وليس مجرد وصف، الأنماط والعلاقات التي قـد توجد داخل مجموعـة البيانات، واختبار قوة وأهمية الارتباطـات بين المتغيرات. كما يشـمل الإحصاء البارامتري (المعلمي) الذي يجري اسـتخدامه لتقييم الفرضيات باستخدام بيانات الفاصل الزمني وبيانات مسـتوى النسبة، مثل الارتباط والانحدار؛ والإحصاءات غير المعلمية المسـتخدمة لاختبار الفرضيات باسـتخدام بيانات المستوى الاسـمية أو الترتيبية؛ والإحصاءات الاحتمالية التي تحدّد احتمال حدوث حالة ما، مثل تحليل النظرية الافتراضية (Bayesian). ويجري تطبيق تقنيات الإحصاء الوصفي والاسـتدلالي التي تستخدم تقليديًا في تحليل البيانات الصغيرة في البيانات الكبيرة أيضًا، وكما جرى نقاشـها في الفصل التاسـع فإن هذا ليس سـهلاً على الدوام لأن الكثير من هذه التقنيات جرى وضعها لرسـم رؤى من البيانات الشـحيحة وليسـت المتوافرة بشكل فائض. ومع ذلك، توفّر هذه التقنيات وسائل للاستفادة من كميات البيانات الهائلة، كما يضمن حجم البيانات الهائل إمكانية تجزئتها إلى مجموعات بطرق عديدة دون أن تصبح صغيرة جدًا مما لا يمكن معه حسـاب الاختلافات بين مجموعتي الضبط والمعالجة بصورة ذات معنى. وتسـتخدم الإحصاءات بشكل روتيني في المساعدة على التنقيب عن البيانات والتنبؤ بها وتحسينها (انظر الجدول ١ - ٦).

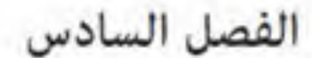
التوقع والمحاكاة والتحسين (Prediction, Simulation and Optimisation):

هنـاك طريقة أساسـية في الحصول على القيمة من البيانات وهي اسـتخدامها لمحاولة معرفة أو توقّع ما سيحدث في ظل ظروف مختلفة. على سبيل المثال، قد ترغب شركة ما في توقـع ردود أفعـال العملاء تجاه منتج معين أو حملة ما، أو قد تحتاج الحكومة المحلية إلى محاولة التنبؤ بكيفية عمل البنى التحتية للنقل إذا جرى إغلاق عنصر مهم ذو حساسية، أو قد يسعى العلماء إلى التنبؤ بموعد احتمال حدوث انهيار أرضي وتحت أية ظروف. إن مثل هـذه المعلومات مفيدة جدا للمنظمات من حيـث تمكينها من التخطيط لحالات الطوارئ المختلفـة، وللشركات من حيث إنتـاج أثر لزيادة نمو الأرباح (انظر الفصل السـابع). وفي جميع الحالات، تُبنى النماذج باستخدام المعرفة القائمة سلفًا عن كيفية عمل النظام، والتي تعالج البيانات لتقدير النتائج المحتملة في ظل سيناريوهات مختلفة. وكما في التنقيب عن

البيانات، هناك العديد من الطرق المختلفة التي يمكن استخدامها لإنتاج النماذج التنبئية. ولكل منها نقاط قوتها وضعفها، وهي تنتج تنبؤات أقل خطأ أو أكثر دقة بالاعتماد على نوع المشكلة والبيانات (Seni and Elder 2010). ومع ذلك، من الصعب الحكم مسبقاً على أن أي نوع من النماذج، وإصداراته المختلفة، سيعمل بشكل أفضل على أيّ مجموعة بيانات مُعطاة. وقد تمثل الحل لهذه المسألة في استخدام نهج مركّب يستفيد من الكمية الهائلة من الطاقة الحاسوبية المتاحة للمحللين الآن (Siegel 2013). فعوضًا عن اختيار نهج واحد وبناء عدد قليل من النماذج، فالنهج المركب يبني نماذج متعددة باستخدام مجموعة متنوعة من التقنيات للتنبؤ بالظاهرة نفسها. وبعد ذلك، بدلاً من تحديد النتائج (التقديرات) من النموذج الأفضل أداء، يجري العمل على الجمع بين التقديرات من جميع النماذج لإنتاج إجابة واحدة شاملة. ويؤدي تجميع النتائج إلى نتائج أكثر قوة كما تعوّض العملية نقاط الضعف في كل نموذج. على سبيل المثال، قد يبني نهجاً مركباً لتوقع سلوك العملاء سلسلة من نماذج الانحدار، والشبكات العصبية، والجار الأقرب، وشجرة القرارات. فكل نموذج قد يكون أفضل من غيره في التنبؤ بأنواع معينة من المستهلكين، ولكن يمكن تقليل الفروقات من خلال الجمع بين مخرجات النموذج، مما يزيد التنبؤ قوة (Franks 2012; Siegel 2013). فباستخدام طريقة تضم العديد من النماذج المنسجمة يمكن تطبيق المئات من الخوارزميات المختلفة على مجموعة بيانات، ومن ثم ضمان إنتاج أفضل نموذج تنبؤي ممكن.

فالمحاكاة هي عبارة عن نماذج تسعى إلى نمذجة العمليات والأنظمة في العالم الحقيقي. والهدف من ذلك هو تحديد كيفية عمل نظام ما وكيفية سلوكه وفقًا لسيناريوهات مختلفة، وتقييم أدائها إحصائيًا بهدف تحسين كفاءتها وفعاليتها (Robinson 2003). ومن الأمثلة الشعبية لعبة الحاسب الآلي سيم سيتي (SimCity) التي تحاكي كيف يمكن لمدينة أن تنمو وتتطور في ظل ظروف يختارها اللاعبون، قائمة على أساس النموذج الكامن وراء العمليات المدنية المعروفة. وبالمثل، تستند توقعات الطقس على محاكاة لكيفية تطور الطقس بالنظر إلى الظروف السائدة والمعرفة العلمية. وهناك العديد من الأنواع المختلفة لنماذج المحاكاة، والتي استفاد كثير منها من التعلّم الآلي من أجل تحسين عمل النموذج بشكل تلقائي والتعامل مع الخصائص الناشئة مثل الأحداث غير المتوقعة. إن نموذج سيم سيتي هو





نموذج قائم على التوكيل (Batty 2007) (Agent-based). ويتكوّن هذا النموذج من بيئة تُمكــن من تعيين خصائص معينــة للميزات الفردية فيها مثل المبــاني والطرق. ويتم تعيين خصائص هذه البيئة بواسطة وكلاء، وعند تشغيل النموذج، يسعى الوكلاء إلى حل مهمة ما، والتفاعل مع البيئة وغيرهم من الوكلاء الآخرين بالاعتماد على خصائصهم المســندة إليهم. وبالمحصّلــة، عند قيام الوكلاء بمهامهم، يقومــون بتغيير البيئة، وهي في هذه الحالة المدينة التي يسكنونها، وإنتاج نظام معقّد ومتغير. ومن ثم فإن النظام يعمل من أسفل إلى أعلى، مع أنماط مكانية وزمانية تتغيّر من تفاعلات الوكلاء الفردية مع البيئة. وتستخدم مثل هذه النــماذج لمحاكاة القضايا المدنية خارج الألعاب لنمذجة اســتخدام الأراضي وتخطيط النقل ووضع خطط الطوارئ للكوارث (Batty 2007).

ويهتمّ التحســين بتحديد الإجراء الأمثل لتحســين الأداء (عادة من خلال خفض التكلفة أو زيادة العائد أو المخرجات). ويمكن حســاب التحســين باســتخدام وتقييم نماذج التنبؤ والمحــاكاة، أو تطويرها من خلال أنواع أخرى من الخوارزميات أو الاختبار الإحصائي. فعلى ســبيل المثال، تســتخدم الخوارزميات الجينية، وهــو نوع معين من التعلّــم الآلي، الأفكار المنتقــاة من الطبيعة، مثــل الوراثة، والطفرات، والاختيار والتهجــين، لوضع وتطوير حلول ممكنة لمشــكلة ما (Mitchell 1996). كما يســعى علم الشبكات العصبية، وهو نهج آخر مســتوحى بيولوجيا، لمحاكاة كيفية عمل الدماغ البشري، باستخدام عناصر المعالجة العالية الترابط لحســاب وتقييم وحلّ مشكلة ما (Picton 2000). كما يمكن استخدام اختبار أ/ب (A/B Testing) بشكل دوري لتقييم وتطويع نظام ما، من خلال مقارنة مجموعة ضوابط مــع مجموعات اختبار مختلفة من أجــل تحديد العلاجات (على ســبيل المثال، النصوص، والتصاميم، والصور، والألوان المســتخدمة على موقع إلكتروني على الإنترنت) لتحسين هدف معين (على سبيل المثال، معدلات التبرع).

الخلاصة:

كان الهدف من هذا الفصل تقديم لمحة عامة عن مجموعة من الآليات الناشئة وتحليلات البيانات التي يتم اســتخدامها لمعالجة وتحليل وتطبيق البيانات الصغيرة الموسّعة والبيانات الكبيرة. لقد تمّ نبذ مثل هذه الأســاليب لعدد من الســنوات، ولكنها أصبحت أكثر انتشارًا



ويمكن الوصول إليها مع تطورها ونضجها، وتوفّر المعالجة الحاسوبية المطلوبة لتصبح متاحة على نطاق أوسـع. ومع ذلك، يطرح حجم وسرعـة البيانات الكبيرة تحديات جدية لآليات التحليل وما قبلها بسـبب عدم تصميم العديد من هـذه الآليات التقليدية والخوارزميات لتتناسـب مع المليارات من الملاحظـات، أو لتتعامل مع مجموعـات بيانات سريعة التغير (Batty et al. 2012). ومـن ثـم فمـن الواضـح أن هناك حاجة إلى مزيـد من البحوث لتكييف، وصقل وتوسيع الآليات الحالية وابتكار أساليب جديدة يمكن لها إضفاء معنى من البيانات الكبيرة والبنى التحتية للبيانات واستخلاص القيمة منها. وعلاوة على ذلك، أضحت الأجيال الجديدة والقديمة من العلماء ومحللي الأعمال، على حد سـواء، بحاجة إلى التعرف عـلى مصادر البيانـات الجديدة وخصائصها، وأن تتعلّم مجموعـة جديدة من المهارات إذا كانت تريد أن تكون قادرة على الدراسة والاستفادة من رؤى وأفكار طوفان البيانات القائم، متضمناً ذلك معالجة البيانات المنظمة والبينات غير المنظمة، والترميز الأساسي، والتنقيب في البيانات، والتمثيل الصوري، والنمذجة، والمحاكاة (انظر الفصل التاسع).

إن هذه التحديات لا تقوم على مجرد أسس تقنية أو موارد بشرية. فإنَّ آليات التحليل تُعدُّ انعكاسًـا لطريقة معينة في الرؤية والاستشـعار للعالم المحيط؛ فهي مظهر من مظاهر المعرفـة الخاصة. وكما نوقش في الفصل الثامن، تثير البيانات الصغيرة الموسـعة والبيانات الكبيرة وآليات التحليل المرتبطة بها تسـاؤلات جوهرية بشـأن تنظيم وممارسـات العلوم (Floridi 2012)، مـع العديد من المقترحـات التي توحي بالدخول في نموذج علمي جديد. ونتيجـة لذلك، لا يتطلب التصدي لطوفان البيانـات مجرد فكر جديد حول طرق التحليل، ولكن تفكيراً متأنيًا فيما يتعلق بفلسفة العلوم (Leonelli 2012). ويستدعي جزء كبير من هذه المهمة التفكير النقدي في التأطير الواسـع للتحليلات وآليات الطرق والمناهج المحددة، والتفكـير في كيفية إنتاج آليات التحليل الجديدة أنواعًـا معينة من المعارف والرؤى للعالم المحيـط بنا، ومـن ثم العمل على معرفة عمـل هذه المعارف الجديـدة في العالم، متضمناً ذلك نمذجة كل الأشـياء التي تشـير إليها (Beer and Burrows 2013; Ruppert 2013؛ انظـر الفصل الحادي عشر). وبعبارة أخرى، هنالـك حاجة ملحة وبصورة عاجلة للبحوث الفنية والفلسـفية فيما يتعلق بآليات التحليل الناشئة من أجل استشعار البيانات الموسّعة والبيانات الكبيرة.


المكتبة الاقتصادية
قناة المكتبة على التيليجرام
ECONLIBRARY





الفصل السابع

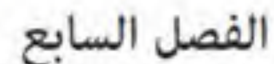
المبررات الحكومية والتجارية للبيانات الكبيرة

(The Governmental and Business Rationale for Big Data)

لم تتكشّف ثورة البيانات دون فكر وبطريقة مجهولة، فكما هو حال الثورات الأخرى، جــرى الدفع بها بســبب مجموعة قوية من الحجج، والتمهيــد لها من قبل مجموعة من المتحمســين المؤمنين بفوائد الطــرق الجديدة للمعرفة والتعامل في هــذا العالم وتحالف أصحــاب المصالح الخاصــة الذين يجنون الأرباح من ظهورهــا. وفي حين أن الحجج التي استخدمت يمكن أن تكون متنوعة جداً، وغالبًا ما تنشأ عن وجهات نظر مختلفة، إلا أنها تشــكّل في مجموعها ما أطلق عليه Foucault (1977) مصطلح نظام الخطابة والمجادلة. أي مجموعة من النقاشــات المتداخلة التي تبرّر وتغــذّي التطورات الجديدة وتعمل على تطبيعها واستنســاخها وإعادة اســتخدامها. فهذا النظام الخطابي يقدّم الأساس المنطقي لتبني الأفكار والتقنيات الجديدة، وإضفاء الشرعية على مساعي تطويرها وسبل تنفيذها. وتدعو النقاشات المستخدمة في النظام إلى الترويج لرسالتهم وجعلها قناعة عامة، لترغيب الناس والمؤسســات باعتنــاق منطقهــم، وإلى الاعتقاد والعمل فيما يتعلــق بهذا المنطق (Dodge and Kitchin 2005). كــما عملــت عــلى إنتاج أجواء معيّنــة يتم من خلالها "تغذية أنواع خاصة من الفكر والممارسة بحيث تبدو طبيعية ومرغوبًا فيها" (Amin and Thrift 2013: 5)، وهي تغرس نوعًا خاصًا من الاســتجابة العاطفية (Anderson 2009). فمثل هذه الخطابات والأجواء ســعت إلى إعــادة صياغة العالم وفق رؤية خاصة، وإعادة تشــكيل "القضايــا، الفاعلين، المخاطر، المصالح، والرغبات التي تنتشر في الحقل الســياسي والمشهد الاقتصادي" (Barnett 2013: 4).

وكما تمّ النقاش في الفصلين الثاني والثالث، فقد سعت نظم الجدال والمحاورة المتداخلة إلى الدفع بعمليات تطوير البنى التحتية للبيانات والبيانات المفتوحة مســتخدمة النقاشات والمداولات حول تبادل ومشــاركة البيانات، واستنســاخها وإعادة استخدامها، وحرية النفاذ





والوصول إليها، والحكومة المفتوحة، والشفافية، والمساءلة، وتنظيم المشاريع الاجتماعية، واقتصاديات السعة ووفورات حجم الإنتاج. وكما تمّ بيانه في هذه الحالات، فإن منحى النقاش وطريقة التداول قد تكون مدفوعة من قبل مجموعة من الأطراف المعنية التي قد تستخدم الحجج نفسها ولكن لتحقيق مرام مختلفة (على سبيل المثال، تستخدم منظمات المجتمع المدني وقطاعات الأعمال الكبيرة خطابات الشفافية والمساءلة لإتاحة بيانات الحكومة وفتحها، ولكن أحدهما لأسباب مدنية والأخرى للانتفاع المالي من البيانات). فالأنظمة الخطابية إذاً لم تكن حصرًا نتاجًا فوقيًا من الدول أو الأعمال، ولكنها انتشرت وسُنّت من خلال العديد من القنوات الأصغر حجماً (Foucault 1977, 1978). وقد جرى قبولها دون إشكالية من قبل الجماهير، من المواطنين والمؤسسات الذين كان من السهل استمالتهم من خلال منطقهم. وبالأحرى، فإن الخطابات مفتوحة للتحدي، والرد عليها، ومقاومتها، وتفنيدها، ومخالفتها، والتصدّي لها، وإظهار عوارها، وتقديم رؤى بديلة جديدة. وفي كلتا الحالتين، يستند المروّجون أو المعارضون لرؤى جديدة على السجلّ الزمني الممتد من الماضي، إلى الحاضر، وحتى المستقبل لدعم حججهم، وذلك باستخدام الأدلة والشواهد من جولات سابقة من التغير الاجتماعي والاقتصادي، إلى جانب التحديات الحاضرة، والتكهنات حول الكيفية التي يمكن أن تشكل الكيفية التي ستكون عليها الأمور خلال سنوات قادمة، في ظل سيناريوهات مختلفة، (Kinsley 2011, 2012). ومن ثم، جرى تداول المشهد الخطابي عبر جولات من التفاوض باسم الصراع على الأفكار والسبل لهيكلة وتمثيل وعمل وتنظيم المجتمعات، في الحاضر والمستقبل.

وفي هذا الفصل، سيجري دراسة المبررات المنطقية للبيانات الكبيرة للجهات الحكومية وقطاع الأعمال. وقد تمّ تنظيم النقاش حول أنماط خاصة من النظم والأنشطة الاجتماعية بدلاً من تنظيم الفصل بحسب مختلف الخطابات التي يشيع توظيفها للترويج لإشاعة اعتماد البيانات الكبيرة على سبيل المثال، التبصّر والحكمة، الإنتاجية، القدرة التنافسية، الكفاءة، الفعالية، الفائدة، الاستدامة، والتوريق المالي (Securitization). فالخطب تمتلك تأثيرها الأقوى عندما يجري تعبئتها وتوجيهها في مهام وقضايا معينة، والمثير في حالة البيانات الكبيرة هو أن جدالها يمتد إلى جميع القطاعات والشرائح - الاجتماعية، والسياسية، والاقتصادية، والبيئية؛ فهي واعدة من خلال تقديمها وسائل جديدة كلياً في فهم

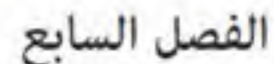


وإدارة جميع جوانب حياة الإنسان. وهنا، سيتمّ مناقشة هذا الوعد فيما يتعلق بأربع مهام رئيسية هي: حكم الناس وتنظيمهم، وإدارة المنظمات، وتعزيز القيمة وإنتاج رأس المال، وأخيراً إنشاء أماكن أفضل. وهذه مهام توضيحية فقط، على سبيل الحصر وليست شاملة، حيث إن هناك العديد من التطبيقات الأخرى الممكنة، بما في ذلك القضايا البيئية (انظر Mayer-Schonberger and Cukier 2013; Siegel 2013; Smolan and Erwitt 2012). وستتم مناقشة كيفية إعادة تشكيل البيانات الكبيرة للبحوث في الفصل التالي، كما سيجري مناقشة الدعوات المضادة لتطبيق البيانات الكبيرة في الفصل العاشر.

حكم الشعب وتنظيمه (Governing People):

الدولة هي مولّدٌ ومستخدم رئيس للبيانات، فمنذ عصر التنوير سعت الدولة لإنشاء طرق أكثر منهجية، من جهة، لإدارة وحكم المجاميع السكانية، ومن ناحية أخرى، لتقديم الخدمات للمواطنين. فإحدى هذه الطرق الرئيسية التي هدفت من خلالها إلى أداء هذه الأدوار بكفاءة وفعالية كانت من خلال التدقيق والتقدير الكمي والنوعي للمجتمع - أي بناء معرفة مفصّلة لإنفاذ ولايتها القضائية والاختصاصية من خلال تطوير النظم الإدارية على نطاق واسع، بدعم من أجهزة فنية وتقنية متنوعة، تعمل على قياس وتتبّع فهرسة وتمييز البيانات عن الأفراد، والشركات، والمؤسسات، والمكونات الأخرى (انظر Alonso and Starr 1987; Barnes and Hannah 2001; Desrosières 1998; Kitchin and Dodge 2011; Porter 1995). إن إنتاج، وإدارة، وتخزين، وتحليل البيانات من المهام الرئيسية في جميع مؤسسات الدولة، فهي تُستخدم في تقييم الالتزامات والاستحقاقات للموضوعات السيادية وغير السيادية، والكشف والمراقبة عن الخروقات وعدم الامتثال، والتهرب، والغش، فيما يتعلق بالمعايير والالتزامات القانونية المعمول بها (Rose 1996). فعلى سبيل المثال، يتمّ في الدوائر الحكومية، وخاصة تلك المتعلقة بالتمويل والتجارة والرعاية الاجتماعية، تحديد جميع المواطنين والكيانات وتتبّعهم، ورصد القضايا والاتجاهات من خلال التحديثات على سجلاتهم، وتعيين ملفاتهم التعريفية وربط بياناتها بهم، وتعقّب المدفوعات، وإدارة انضباطهم والخدمات المقدمة لهم. أما في المدارس، فيؤخذ التسجيل لبياناتهم، واحتساب وتقييد الدورات الدراسية وعلامات الاختبارات، ويتمّ مكافأة الطلاب





ومعاقبتهم وفقًا لأدائهم وسلوكهم. وفي قطاع الصحة، يتم قياس حالة المريض، وتسجيلها، والعمل على أساسها، واحتساب وتقييم أداء المرافق الصحية. كما تُعنى وكالات حكومية متخصصة أخرى بإنتاج وإدارة وتحليل ونشر البيانات التي لها بالغ الأثر في إفادة العمل الحكومي. على سبيل المثال، الوكالات الإحصائية، وهيئات المسح ورسم الخرائط الجغرافية، وأجهزة الاستخبارات التي تتعهّد ببرامج مستمرة لإنتاج البيانات من خلال أنظمة قياس مثل المسح والمراقبة.

ولتعزيز ممارساتها الحسابية، سعت الدول، من جهة، لتبرير وترشيد نظمها وعملياتها على الموضوعات، ومن جهة أخرى، لتطبيق وسائل إدارية أكثر تعقيدًا وتفصيلاً من أي وقت مضى (McNay 1994; Miller 2001). فالمسعى الأول يشمل الخطابات حول الحكومة الفعالة والقيمة مقابل المال، والمواطنة والتمكين، والعدالة ومكافحة الجريمة، والأمن والسلامة، والتي تتماشى مع فكرة تناول المشاكل والقضايا العامة من خلال المنطق الجماعي والمبادئ المشتركة. وتُجابه هذه الخطابات من قبل أولئك الذين لديهم الرغبة في تقليل مستوى رقابة الدولة وسلطتها مقابل الحريات الشخصية، والرقابة والضبط للنظم المجتمعية والاقتصادية، أو أن تكون أكثر انفتاحًا واستجابة وإنتاجية في الممارسات الحكومية، متضمناً ذلك إتاحة البيانات التي يحملونها بشكل أكبر وأكثر حرية (انظر الفصل الثالث).

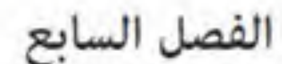
أما المسعى الثاني فيتضمن تنفيذ أنظمة وتقنيات وبنى تحتية مصممة لجمع ومعالجة البيانات على نحو فعال وأكثر تطورًا من أي وقت مضى، متضمناً ذلك الأنظمة الأكثر ثباتًا وقدرة في التعريف والتحديد (التواقيع والصور وبصمات الأصابع، وأرقام الضمان الاجتماعي وجواز السفر، وقوانين الضرائب، وأرقام الترخيص والحسابات، والمعلومات البيومترية، والمستخدمين وكلمات السر والرموز البريدية... إلخ)، والتي يمكن تشغيلها على نحو متزامن بشكل متزايد وعن بُعد (Dodge and Kitchin 2005). وعليه، يؤكّد (178 :1997) Hannah على أن أحد أشكال الحاكمية الأكثر فعالية يكون في الغالب من خلال تكوين "لفيف معقد من مجموعة بسط النفوذ والهيمنة...من خلال التعدادات والبيروقراطية والسياسات، وما إلى ذلك، وكم هائل من المؤسسات التخصصية مثل المدارس وأماكن العمل، والأراضي الممتدة المجهولة المثبتة بملايين من نقاط الانطلاق كأجهزة الصراف الآلي، وكاميرات المراقبة، ونماذج تعبئة الطلب".



ولذا، فإن البيانات الكبيرة بأحد معانيها ليست شيئًا جديدًا على الدول - وهي التي تعاملت طويلاً مع كميات هائلة من البيانات في إدارة الاختصاصات القضائية والقانونية، كجزء جرى معالجته واتخاذ الإجراءات المتعلقة به في الوقت المناسب في حينه أو قريبًا منه كلما تفاعلت الوكالات مع الموضوعات. ومع ذلك، يُنظر إلى البيانات الكبيرة باعتبارها أحدث مجموعة من التقنيات التي يمكنها توسيع وتحسين عمل الدول من خلال زيادة الانضباطية والشمولية للممارسات الحسابية لأعمالها. كما يتبنى البائعون، في بعض الحالات، خطاب الدفع بحلول البيانات الكبيرة بهدف بيع خدمات الدولة، وفي حالات أخرى، تأخذ الدولة بزمام المبادرة بشكل نشط. وفي الوقت الحاضر، هناك نوعان من الإستراتيجيات الخطابية الرئيسية المستخدمة لتشجيع استخدام البيانات الكبيرة في السلطات الاختصاصية، يتعلق النوع الأول بالتحسينات على الإدارة العامة وتقليل التكاليف من خلال تعزيز الكفاءة التشغيلية، وتقليل تكلفة الأخطاء والاحتيال في إدارة المنافع، وزيادة في الإيرادات الضريبية عن طريق تضييق الفجوة الضريبية... وتحسين تخصيص التمويل في البرامج والخدمات ذات الجودة الأعلى، وزيادة المساءلة في القطاع العام، وإطلاع أفضل للمواطنين"، مع تقديرات مقترحة بعائدات لحلول البيانات الكبيرة من الكفاءة التشغيلية المتعلقة بإدارة القطاع العام في القارة الأوروبية تعادل مئة وخمسين مليار دولار أمريكي (Manyika et al. 2011: 61, 62). وسيجري لاحقًا مناقشة هذه الحجج بشكل أكبر فيما يتعلق بإدارة المنظمات وصنع أماكن أفضل.

ويهتمّ النوع الثاني بأمن الدولة ومكافحة الجريمة، إذ تقوم جميع الدول بالمراقبة والرصد لأغراض الأمن والسلامة ومنع الجريمة والاعتقال من خلال أجهزة الشرطة وجمع المعلومات الاستخباراتية على نطاق أوسع. وقد تمّ تطوير الكيفية التي تنفّذ بها هذه الأنشطة بما يتماشى مع التقنيات الجديدة - كالتنصت على المكالمات الهاتفية واعتراض الموجات الراديوية، والدوائر التلفزيونية المغلقة، والتصوير الجوي، وصور الأقمار الصناعية، ومراقبة الإنترنت - والتي كثيراً ما جرى التبرير لها عند وقوع أحداث رئيسية مثل الحرب الباردة، والحوادث الإرهابية (Innes 2001). وفي الآونة الأخيرة، قدّمت أحداث الحادي عشر من سبتمبر وما تبعها من الحرب على الإرهاب المبرّر المنطقي لتغيير جذري في عمليات المراقبة والتأمين التي تقودها الدولة على نطاق واسع (Amoore 2006). وكان جليًا لبعض





الوقت أن تحليلات البيانات الكبيرة كانت عنصرًا أساسـيًا في تحسين هذه الأنشطة، وخاصة فيــما يتعلق بمراقبة الحــدود وفحص الهجــرة (Kitchin and Dodge 2006; Clements 2013). وقد كشــفت تسريبات يونيو 2013 بشــأن برنامج وكالــة الأمن القومي الأمريكي بريزم (PRISM) المدى الذي وصلت إليه ممارســات مراقبة البيانات، وممارســات البيانات الكبيرة، والآليات التحليلية المستخدمة لجمع ومعالجة البيانات المتعلقة بالاتصالات اليومية لجميع المواطنين مثل المكالمات اليومية، الرسائل النصية، رسائل البريد الإلكتروني، واستخدام الإنترنــت، بدلاً من اقتصارها على المواطنين الأجانــب والقوى الخارجية أو لأهداف محددة (Greenwald and MacAskill 2013; Risen and Lichtblau 2013). فهذه التسريبات كانت هي الأحدث في سلســلة الكشف عن تجسس الولايات المتحدة على مواطنيها بشكل خفي (Solove 2007 ; Bamford 2009). والواقع، أن من الواضح أن الوكالات الاستخبارتية، والمتعاقدين العسكريين-الصناعيين، هم في طليعة مطوري الجيل القادم من أدوات البيانات الكبــيرة لإدراك الكميــات الهائلة من البيانات عالية السرعة، وقد اســتثمرت بكثافة في بناء الكفاءات والتقنيات والمرافق على ســبيل المثال، جرى اســتثمار 1.7 مليار دولار أمريكي في مركــز البيانات القائم في يوتا لمعالجة وتخزين البيانات الكبيرة التي تم جمعها من الولايات (Carroll 2013). وبينما تســعى الدول إلى ترشــيد تلك المراقبة ذات النطاق الواسع على مواطنيهــا من خلال الخطابات العامة الموجهة للأمن والســلامة، فــإن الحقيقة الماثلة هي أنهــا قد طوّعت الكثير من أجهزتهــا سراً لتنفيذها ودون اتباع للإجراءات القانونية الواجب العمل بهــا مما يقوّض المثل العليا للديمقراطية القائمة على الشــفافية والثقة (Clements 2013). ومع ذلك، فمن غير المرجح على المدى القصير أن يقوم الغضب الشعبي العارم على هذه الممارســات والتجاهل الصارخ للحقوق الأساسية مثل الخصوصية، على إبطاء أو إعادة تشكيل طريقة التعامل مع البيانات الكبيرة في هذا الجانب من الحكم، ولا لثني الخطابات التي استخدمت لتبريرها.

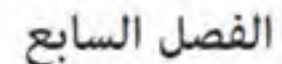


إدارة المنظمات (Managing Organisations):

تمامًا كما تغيرت الدول وتفرعاتها الكثيرة في الحكم كالإدارات، السلطات المحلية، الوكالات، الجيش، مقدمي الرعاية الصحية... إلخ مع مرور الوقت، واعتمادها ممارسات وتقنيات إدارية جديدة لإعادة تشكيل الكيفية التي يجري بها تنظيمها وتشغيلها، بالمثل، قامت شركات الأعمال والكيانات المؤسسية الأخرى، مثل المنظمات غير الحكومية والهيئات التطوعية والمجتمعية، والجمعيات الخيرية، والأحزاب السياسية... إلخ. فالبيانات المتعلقة بمهامها الأساسية تشكّل أحد الجوانب الأساسية لكيفية فهم هذه المنظمات لنفسها ودوائرها الانتخابية، وكيفية اتخاذ القرارات بشأن عملياتها الحالية والمستقبلية. كما أن البيانات تقدّم الإثبات والدليل للإنتاج المعرفي الذي يستخدم لإدارة المنظمة بشكل أكثر كفاءة، وفعالية، وتنافسية، وإنتاجية، وتحديد المنتجات والأسواق الجديدة (انظر القسم التالي).

وهكذا، أصبحت نظم المعلومات على مدى السنوات الستين الماضية من البنى التحتية الأساسية لجميع المنظمات على اختلاف أحجامها، مما يمكّنها من تتبّع وإدارة المجاميع البشرية المعقّدة كالموظفين، الداعمين، المانحين، والعملاء، والمكونات، والسلع الأساسية، والبنى التحية عبر الزمان والمكان، وتمكينها من توسيع نطاق عملياتها من المستوى المحلى إلى المستوى العالمي (Castells 1996; Dicken 2003). فنظم المعلومات تتيح تشغيل المنظمات بشكل أكثر ذكاء، وإنتاج وفورات كبيرة من تكاليف قاعدتها التشغيلية. وهناك ثلاثة من هذه النظم المعلوماتية المبنية على البيانات والتي تمّ اعتمادها على نطاق واسع: نظم تخطيط موارد المنظمات (ERP)، ونظم إدارة سلسلة التوريد (SCM)، ونظم إدارة علاقات العملاء (CRM). وتسهّل هذه الأنظمة قدرًا أكبر من التنسيق والضبط داخل المؤسسة، ومع المنظمات الأخرى، والعملاء. على سبيل المثال، ينظّم ويجمع نظام تخطيط موارد المنظمات العديد من قواعد البيانات والأنظمة المتعلقة بالشراء، التخزين، الجرد، النقل، التسويق، المحاسبة، إدارة شؤون الموظفين وإعداد قوائم المرشحين، إدارة المشاريع، والعلاقات مع العملاء في نظام واحد شامل يضمن بصورة سهلة وتلقائية توافر البيانات والعمليات من جزء واحد من الأعمال لجزء آخر (Dery et al. 2006; Grant et al. 2006). كما تُستخدم نظم إدارة سلسلة التوريد (SCM) بأكبر قدر من الكفاءة لتنظيم المشتريات، والحركة، وإدارة وتخزين المواد من الموردين من الشركة وحتى





الزبائن (Chopra and Meindl 2012). في حين تسـعى نظم إدارة علاقات العملاء (CRM) إلى بناء علاقات شـخصية مع العملاء من خلال جمعها أنماط الحجوزات والمشـتريات السابقة لهم، وعرض التفضيلات والخدمات الممكن الاسـتفادة منها خلال زياراتهم، والاستفادة منها في تقديـم مكافآت الولاء، والعروض الخاصة، والترقيـات، وتقديم التوجيهات للموظفين والأنظمة ليظهروا بمظهر العارف بالعميل (Sigala 2005). وقد أكّد Brynjolfsson et al. (2011) من خلال دراسـة شملت 179 شركة من شركات التداول العام الكبيرة على اعتماد هذه الشركات في اتخـاذ القرارات المبينة على البيانـات (Data Driven Decision- DDD) حيث يكون الناتج والإنتاجيـة أعـلى بما يُراوح بـين ٥-٦ ٪ عما هو متوقع عند الاعتماد على اسـتثماراتهم الأخرى وطرق استخدام تقنية المعلومات.

ويُنظـر إلى البيانات الكبـيرة من قبل العديد على أنها المرحلة التاليـة في العملية الجارية التي تسـعى إلى الكفاءة التنظيمية والبنى التي تعزّز الميزة التنافسـية، مع الحد من المخاطر، والتكاليف، والخسـائر التشغيلية، وتحسـين تجربة العملاء. فالذي توفره البيانات الكبيرة هو الإمكانية الشـاملة والمفصّلة لغاية إدراك جميع جوانب عمل المؤسسـة بشـكل آني ولحظي. وكـما افترضت جماعة الضغط، مؤسسـة تك أمريـكا (TechAmerica Foundation 2012: 12)، فإن البيانات الكبيرة سـتضع الأسـاس: لاسـتبدال أو دعم عمليات اتخاذ القرار البشرية بخوارزميات آلية، مما يعمل على الحد من أوجه القصور داخل المنظمات؛ ويوجد الشـفافية؛ ويُحسن الأداء من خلال تمكين التجريب لاكتشاف الحاجات وإظهار التباينات؛ وتحسين العائد عن الاسـتثمار (ROI) لاستثمارات تقنية المعلومات، وتحسـين عملية اتخاذ القرار والتشغيل الذكي؛ وتوفير قدرات تنبؤية لتحسـين مخرجات المهام؛ والحد من التهديدات الأمنية والجريمة؛ والقضاء على الهدر، والاحتيال، وإسـاءة الاسـتخدام والمعاملة؛ وابتكار نماذج أعمال وخدمات مستفيدين جديدة.

وبعبـارة أخرى، سـتوفر البيانات الكبـيرة معرفة لحظية غاية في الدقـة المفصّلة لكبار المديرين في جميع مناحي منظماتهم والتي يمكنهم من خلالها صياغة دورات عمل مُحكمة. وسـتمكّن البيانات الكبيرة المنظمات مـن أن تكون أكثر فطنة، وملائمـة، ومرنة، ومبتكرة، وذكية في كيفية تنظيمها وتشغيلها (Manyika et al. 2011).

ومثال علـى كيفية التوظيف الناجـح للبيانات الكبيرة من قبـل منظمة غير تجارية لإعـادة هيكلة وإدارة عملياتها الحملـة الانتخابية لباراك أوباما عن الحزب الديمقراطي. فكـما أشرنا في الفصل الرابع، فقد تبنّت الحملة الانتخابيـة لأوباما البيانات الكبيرة، من خلال جمع مجموعة هائلة من البيانات من مصادر متعددة وإنتاج بيانات اقتراع معتبرة لحظياً في الوقت المناسـب، من أجل إدارة ردّ الفعل، وتشكيل الإستراتيجية، والإجراءات، ومهام العاملين والمتطوعين. واسـتنادًا إلى تحليل الأداء عـام ٢٠٠٨، فقد نمت مجموعة تحليلات البيانات للحملة الانتخابية لأوباما للعام ٢٠١٢ في الحجم خمسـة أضعاف عنها في حملة ٢٠٠٨ متضمنة روّادًا تقنيين جرى التعاقد معهم من سوق الصناعة (Scherer 2012). وقد حسّـنت هذه المجموعـة العلائقية بين البيانات التي تمّ جمعهـا من مصادر مختلفة بحيث أمكن ربطها معاً بشكل أكثر فعالية. كما طوّروا تطبيقات للحملة الانتخابية، وجرى استخدام شـبكات التواصل الاجتماعي مثل فيسـبوك للدفع باتجاه التسجيل للانتخاب والحصول على التصويت لصالحها، وتثبيت ملفات تعريف الارتباط (Cookies) الخاصة بهـم وتلك المتعلقة بأطراف تـم التعاقد معها على أجهزة الزائريـن لموقعهم الإلكتروني لجمـع البيانات عنهـم وتتبّع عاداتهم عـلى الشـبكة (Crovitz 2012; Kaye 2012). كما قاموا بتحسـين تصنيفاتهـم التعريفية وتوقعات النماذج التنبئية وكيفية اسـتخدام المعلومات الناتجة عن تحليلاتهم لتوجيه الحملة، بالإضافة إلى اختبار وصقل طرق زيادة التمويـل لدعم الحملة ماليًا (Scherer 2012). كـما أضافوا وعالجوا البيانات الجديدة بشكل مسـتمر مع تشـغيل لأنظمة المحاكاة للتنبؤ بالنتائج وأفضل الردود. وكما صرّح أحد مسؤولي الحملة: "لقد شغّلنا نموذج المحاكاة ٦٦,٠٠٠ ألف مرة في كل ليلة" لتحديد احتمالات الفوز في كل ولاية متأرجحة". وفي كل صباح يصل إلينا الناتج، هذه هي فرص نجاحكـم في الفـوز في هذه الولايات، وبهذه الطريقة تـمّ تخصيص الموارد (مقتبس عن Scherer 2012). لقـد حصـل فريـق أوباما على مزايا مهمة عن منافسـيهم من خلال التشـغيل المسـتمر لمجموعات قواعد البيانات المتغيرة من خـلال خوارزميات متطورة، وعليه أشـار Issenberg (2012: 246) إلى أن حملات أوباما الانتخابية كانت "المجموعة السياسـية المثالية: المؤسسـية المموّلة جيّداً، القائمة على البيانات، والممارسـة التجريبية الصارمة "، وهذا هو وعد الدعاة للبيانات الكبيرة بأن الفوائد العائدة عنها كالتي اكتسبها



الحــزب الدميوقراطي من البيانات الكبيرة، لها أن تتحقــق لجميع المنظمات وغيرها من شركات قطاع الأعمال.

تحصيل القيمة وإنتاج رأس المال (Leveraging Value and Producing Capital):

فضــلاً عن جعل الــشركات أكثر كفاءة وفعالية في إدارتها، يتــمّ الترويج بكثافة للبيانات الكبيرة لتكون وسيلة لرفع مستوى القيمة الناتجة من أنشطة الشركة وعلاقاتها مع الموردين والمســتهلكين. وفي الواقع، وكما أشــار (2011) Gantz and Reinsel، فإن المحرك الرئيسي للبيانــات الكبيرة ليس تقنيًا؛ ولكنه ماليٌّ مع الوعد بالمزيــد من الكفاءة والأرباح. ومن ثم فــإن اعتماد حلول البيانات الكبيرة هو إســتراتيجية لإدراك تراكم رأس المال غير المســتغل بطريقــة تقدّم عائداً كبيراً من الاســتثمار وتوفير ميزة تنافســية (Manyika et al. 2011; Zikopoulos et al. 2012). وهنــاك العديــد مــن الطرق التي مكــن من خلالها لحلول البيانات الكبيرة أن تُضفي الذكاء للشركات وبطريقة مكن معها أن تنمو مبيعاتها وأرباحها. فمــع الأخذ لقطاع البيــع بالتجزئة كمثال، حدّد Manyika et al. (2011) ســتة عشر حلاً مقسّمة إلى خمسة أنواع - التسويق، والترويج، والعمليات، وسلسة التوريد، ونماذج الأعمال الجديــدة - مكن لكل منها تحصيل مكاســب محتملة في مجــال المبيعات والكفاءة (انظر الجــدول ١-٧). وهي تشــمل إنتاج بيانــات عبر الشركة وتطبيق آليــات التحليل من أجل تقســيم الســوق، والتصدي لعزوف العملاء والموظفين، وتحســين المدخلات المختلفة (على ســبيل المثــال، المكونات، والعمالــة، والمرافق) وتؤدي من ثم إلى بناء مختلف التشــكيلات والنماذج التنبئية للإجابة عن مجموعة متنوعة من الأسئلة من قبيل:

- هل يتم الاتصال بالعميل أم لا (التسويق المستهدف)؟
- هل يتم تقديم عروض للاحتفاظ بالعميل أم لا (المحافظة على العملاء)؟
- أي نوع من الإعلان أو اختيار الكلمات / الصور أو المنتج تقدم للعميل (انتقاء المحتوى)؟
- أي قناة ينبغي الاتصال من خلالها بالعميل (اختيار القناة)؟
- هل يتم تقديم عرض أعلى أو أقل سعرًا للعميل (التسعير/الخصم المتغير)؟
- هل يتمّ شطب جزء أكبر من الدين عن المقترض (التحصيل)؟



• هــل يتم تقديم حــد ائتماني أعلى أو أدنى أو معدل فائــدة للعميل (مخاطر الائتمان)؟ (Siegel 2013).

وللتركيز على نوع واحد من أنواع الحلول الخمسة التي قدّمها Manyika et al. (2011)، وهو التســويق، يمكن اســتخدام عدد من آليــات التحليــل وأدوات إدارة البيانات الكبيرة الموزعة في أربعة محاور هي: التوحيد والاختزال، التقســيم والتجزئة، المطابقة، والتحســين، وذلك لإدارة ومعالجة وتحليل وتفسير البيانات من أجل تلبية الأهداف (انظر شكل ١ - ٧). إن هدف التسويق السلوكي هو بيع البضائع والخدمات بصورة أكثر فعالية للأفراد والعملاء وبناء علاقة طويلة المدى معهم من أجل الحفاظ على عاداتهم وذلك من خلال الانسجام مع خصائصهم، وســلوكهم، وميولهم، ورغباتهم (Schwartz and Solove 2011). فمن خلال بناء الملفات التعريفية الفردية للأشــخاص ودمجها مع التحليلات التنبئية، تسعى الشركات إلى تقديم تجربة شــخصية من خلال التســويق الجزئي للمنتجات والعروض الخاصة. وفي ســبيل ذلك، تســعى الشركات إلى توقّع نية العميل في الشراء وتسهيل تحقيقهم لهذه النية بلطف، مع الســعي لبناء ولاء العملاء في الوقت نفســه (Minelli et al. 2013). وفي ذات المعنى، زعم Bill Stensrud أن "الموزعين يســتخدمون البيانات الكبيرة في محاولة للحصول على المستهلكين" (مقتبس في Bollier 2010: 24).

وقــد تمّ بالفعل تنفيذ مثل هذا النهــج للبيانات الكبيرة من قبل عدد من تجار التجزئة مثل وول مارت وتيســكو. حيث تنتج مثل هذه السلاســل من المتاجــر كميات هائلة من البيانــات يوميًا من خلال نظــم تخطيط موارد المنظمات (ERP)، وإدارة سلســلة التوريد (SCM)، وإدارة علاقــات العمــلاء (CRM) الخاصة بها، بما في ذلك ربــط منافذ المبيعات بالزبائــن من خــلال بطاقات الولاء والائتــمان (في حالة وول مارت هنــاك ما يقارب ٢٦٧ مليون معاملة يومياً)، واستخدامها لوضع إستراتيجيات أفضل للأسعار، ومخططات التخزين، وإعــداد قوائم الموظفين، ومراقبة المخزون، والحملات الإعلانيــة، وتكاليف المورد، وتحديد مواقــع جديدة للتخزين والمنتجات،...إلخ، واقتطاع حصة من ســوق منافســيها عند القيام بذلك (Bollier 2010). وعوضًا عن إجراء مسوح عينية ضيقة ودراسات استقصائية باهظة التكلفة، لاستطلاع آراء المشاركين عن ماذا سيفعلون، وما هي القيمة المعتبرة بنظرهم، تقوم

البيانــات الناتجة عن العمليات والتصفح على الشــبكة بالكشــف عن هذه المعلومات عبر جميع المســتهلكين. فالفوائد المتوقعة مرتفعة، حيث يشــير Manyika et al. (2011) إلى أن البيانات الكبيرة يمكن أن تزيد هوامش التشــغيل لتجار التجزئة بنسبة ٦٠ في المائة، وفي حين أن هذه الحلول المتعلقة بمبيعات التجزئة والتســويق من الممكن أن تتوفر في مختلف القطاعات، فإن درجة بروزها يختلف من قطاع لآخر بين مختلف القطاعات الأكثر استفادة من حلول معينة (انظر الجدول ٢-٧).

الجدول رقم (١ - ٧)

أنماط ذكاء البيانات الكبيرة للشركات

التسويق (Marketing)	البيع العابر للحدود (باستخدام المواقع الجغرافية للزبائن، وتاريخ الشراء، والتفضيلات، والمعطيات الأخرى لزيادة متوسط حجم الشراء)
	التسويق القائم على الموقع (باستخدام الموقع الحالي بشكل آني أو لحظي، على سبيل المثال نظام تحديد المواقع الجغرافية في الهواتف الذكية، واستهداف الزبائن الموجودين في المتاجر أو بالقرب منها)
	تحليل السلوك داخل المتاجر (تتبّع الهواتف الذكية، وسلال التسوّق، والمراقبة المرئية للمساعدة في تحسين مخطط توزيع المتجر، وتنوّع المنتجات، ومواضع الأرفف لزيادة المبيعات)
	التقسيم الجزئي للعملاء (استخدام تاريخ الشراء والتصفح لاستهداف العملاء بشكل فردي بالإعلانات والعروض الخاصة)
	التحليل العاطفي (تتبّع ورصد آراء العملاء على مواقع التواصل الاجتماعي لتحديد ميلهم تجاه منتج ما وتعديل الحملات التسويقية بشكل مناسب)
	تعزيز تجربة العملاء مع تعدد القنوات (إنشاء تجارب وفهارس على شبكة الإنترنت مصممة بشكل فردي بحسب العملاء)



التّبضّع (Merchandising)	تحسين التشكيلة (تخصيص البضائع لحفظها في أيّ من المتاجر استنادًا إلى المواقع الجغرافية المحلية، وانطباع المشترين وسلوكهم، والحقائق المحلية الأخرى) تحسين التسعير (باستخدام السلوكيات الشرائية والجغرافية لاتخاذ قرارات التسعير في مختلف المواقع لزيادة المبيعات / الأرباح) تحسين التموضع والتصميم (باستخدام السلوك داخل المتجر والتفاعلات على الموقع الإلكتروني لتوجيه عملية تصميم الموقع / المتجر وأماكن وضع البضائع لزيادة المبيعات)
العمليات (Operations)	شفافية الأداء (التحليل المفصّل، بشكل آني أو لحظي على نحو متزايد، لتحديد مدى أداء المنتجات، والمتاجر، والموظفين) تحسين مدخلات الموظفين (احتساب المستويات المثلى للتوظيف وجدولة المناوبات لتفادي فرط التوظيف أوشح القوى العاملة)
سلسلة التوريد (Supply Chain)	إدارة المخزون (تحسين توقع المخزون من خلال رصد تواريخ المبيعات، ودورات المبيعات الفصلية، والأحداث والظروف المحلية لضمان مستويات مناسبة من المخزون المطلوب) تحسين التوزيع والدعم اللوجستي (تحديد مواقع البضائع بشكل آني أو لحظي، واحتساب مسارات الطرق الأكثر ملاءمة وكفاءة من حيث الوقت والوقود، ومراقبة سلوك السائقين، وجدولة أعمال الصيانة) مفاوضات الموردين (باستخدام البيانات المتعلقة بتفضيلات المستهلكين ومشترياتهم لإبلاغ مزودي المنتجات)
نماذج الأعمال الجديدة (New Business Models)	خدمات مقارنة الأسعار (تقديم خدمات مقارنات شفافة من أطراف ثالثة للزبائن لأسعار المنتجات بشكل آني أو لحظي عبر مختلف تجار التجزئة) الأسواق القائمة على الويب (تجار التجزئة الذين يقدمون خدمات بقيمة مضافة مثل الوصول إلى الباعة الآخرين، ومعلومات المنتجات، والمراجعات الناتجة عن المستهلكين، ومقارنات الأسعار)

المصدر : جُمعت من Manyika et al.2011:67-71.



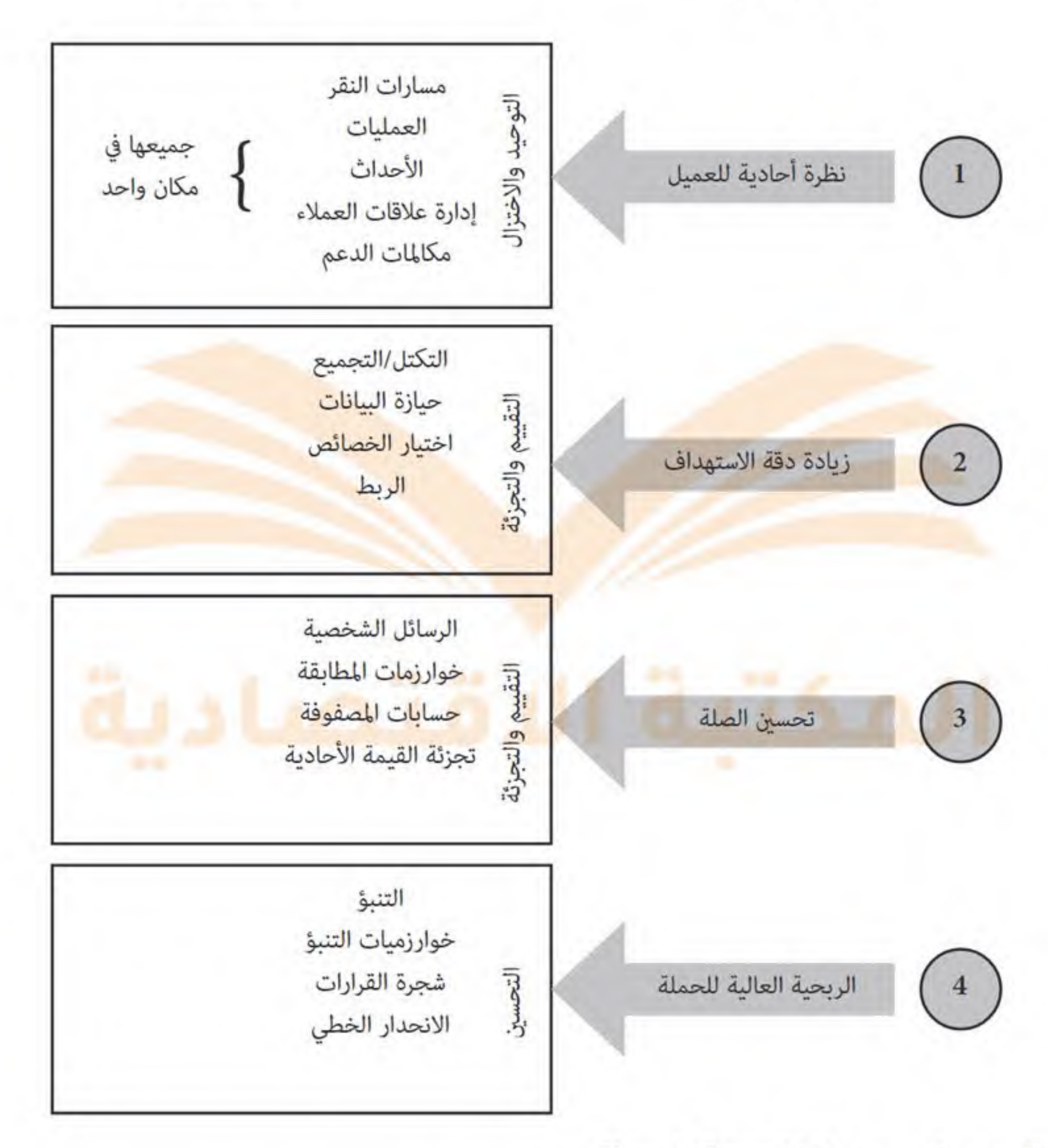
مقتبـــس من "البيانـــات الكبيرة: آفاق جديـــدة للابتكار والمنافســـة والإنتاجية"، (معهد ماكينزي العالمي Mckinsey Global Institute (May 2011)

.(www.mckinsey.com/mgi.Mckinsey&Company)

وإضافة إلى الدفع بالفوائد المتوقعة، سعى دعاة البيانات الكبيرة أيضًا إلى دعم اعتمادها من خلال تسليط الضوء على عواقب فشل الاستثمار فيها. على سبيل المثال، أشار Manyika et al. (2011: 6) إلى أنهـــم: "يتوقعون أن يـــرون، عبر القطاعات، تحقق القيمة التي تعود على المســـتخدمين الرئيسيين للبيانات الكبيرة على حساب المتقاعسين عنها"، وبعبارة أخرى، سوف يتخلف أولئك الذين لا يستطيعون التأقلم واعتماد حلول البيانات الكبيرة عن الركب ويتعرضون للتعثر والأفول؛ لأنهم ســـيفقدون حصتهم في السوق. وبغض النظر عمًّا إذا كان اســـتخدام الخطابات للدفع أو الجذب، فقد أضحت البيانات الكبيرة للأعمال منتشرة داخل التيار الرئيسي ووسائل الإعلام التجارية منذ العام 2012م، وبالنظر لفوائدها الموصوفة، يبدو أن البدء في تطبيقها على نطاق واســـع ســـيمثل المرحلة التالية للابتكار التقني وذلك بصورة مؤكدة.

الشكل رقم (١ - ٧)

التسويق والبيانات الكبيرة



المصدر: مقتبس من Claverie-Berge 2012.

تم إعادة الطباعة بإذن من المؤسسة الدولية لآلات الأعمال، © المؤسسة الدولية لآلات الأعمال (International Business Machines Corporation).



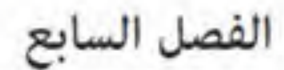
الجدول رقم (٢ - ٧)

فوائد البيانات الكبيرة لعشر صناعات مختارة

صناعة السيارات	التأمين
المراقبة المتقدمة للحالة تجميع وتسعير السيارات الجديدة مطالبات الضمان	تحديد التصرف القادم الأمثل نمذجة الكوارث الغش في المطالبات
الخدمات المصرفية	**النفط والغاز**
تحديد التصرف القادم الأمثل تحليل سجل النظام إدارة مخاطر السيولة	المراقبة المتقدمة للحالة مراقبة وتحسين الحفر مراقبة وتحسين الإنتاج
الطاقة والخدمات	**التجزئة**
تحليلات العداد الذكي التنبؤ بتوزيع الأحمال وجدولتها الصيانة بحسب الحالة	التحسين الآلي للأسعار التشكيل المحلي تحسين التسويق
الحكومة	**الاتصالات**
منع الجرائم والتهديدات والتنبؤ بها تحليلات الاحتيال والنفايات وإساءة المعاملة	ضمان الخدمة الأفضل ذكاء الشبكة استحداث منتجات وخدمات جديدة
الرعاية الصحية	**السفر والنقل**
التدبير والتصرف بناء على الصحة العامة للمجتمع إشراك المستهلكين في رعايتهم الصحية	تحليلات العملاء وولائهم التسويق الصيانة التنبئية

المصدر: أي بي إم (IBM (n.d.))

http://www-01.ibm.com/software/data/bigdata/industry.html





إيجاد أماكن أفضل:

اخــترع الناس واعتمدوا على مر التاريخ على تقنيات جديدة صممت لإنشــاء الثروة وتحســين الحياة اليومية وتعزيزها. وقد أثرت العديد من هذه التقنيات بشــكل كبير على طبيعة وهيكل وتنظيم وسنّ الحياة المدنية والريفية، في الخدمات الخاصة المُقدمة، والنقــل، والاتصــالات، وبناء البنية التحتية. فهذه التقنيــات تحمل في كثير من الأحيان تأثيرات متناقضة وســلبية على ســبيل المثال، المصانع التي صنعت الثروات، وفي الوقت نفســه التلوث البيئي أيضًا، أو أن تكون مبالغاً فيها بحيث تتجاوز الطاقة الاســتيعابية على ســبيل المثال، الاكتظاظ أو الازدحــام المروري، أو تنتج التبعيــات التي لا يُمكن أن تُطاق على سبيل المثال، المسافات الطويلة بين مواقع السكن ومواقع العمل. إن تقنيات المعلومات والاتصالات (تكنولوجيا المعلومات والاتصالات) هي أحدث موجة من الأجهزة والبنــى التحتية الجديــدة التي أضحت جزءًا لا يتجزأ في نســيج البيئة التي نعيش بها، ومعها مكوّن مهمٌ من مكونات منطقها الاســتطرادي وهو أنها تعمل على توفير الحلول للمشاكل الناجمة عن الدورات السابقة من التنمية التي قادتها التقنية – بحيث تجعل منها أكثر كفاءة واستدامة، ونظافة، وأمناً، وإنتاجية... إلخ – وكذلك إنشاء طرق جديدة تمامًــا للعمل في العالم المحيط. ونتيجة لذلك، تبنّت المُدن تقنية المعلومات والاتصالات، على مــدى العقود القليلة الماضية، باعتبارها عنصرًا أساســيًا في إســتراتيجياتها الإنمائية (Castells 1996; Graham and Marvin 1996; Kitchin 1998).

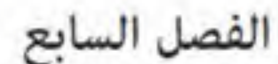
إن مفهــوم المدن الذكية هو تجســيد لهــذه الرؤية التــي تقودها تقنيــات المعلومات والاتصالات للتنمية المدنية، والذي يصف الأماكن التي تتكون بشكل متزايد وتُراقب بواسطة الحوسبة المنتشرة والحوسبة في كل مكان، ويقود ضبط اقتصادها وإدارتها الإبداع، والابتكار، وريــادة الأعمال، من قبل أنــاس أذكياء. وهذه المدن الذكية مجهزة بشــكل مكثّف ويمكن فهمها وضبطها بشــكل آني أو لحظي؛ فهي تنتج، وتتبادل، وتدمج، وتســتهلك، وتعمل على البيانات الكبيرة (Kitchin 2014). إن مثل هذه البيانات الكبيرة توفّر فهماً شاملاً، ومتغيراً، ومفصّلاً بدقة لهذه المدن والمدخلات للأنظمة التي من شأنها بناء أماكن أكثر ملاءمة للعيش، وآمنة، ووظيفية، وتنافسية، ومستدامة (Hancke et al. 2013; Townsend 2013). فمثل



هــذه الرؤيــة للمدن الذكية يجري الترويج لها بشــكل كبير من قبل عــدد من أكبر مزودي خدمات البرمجيات في العالم وشركات الأجهزة الرقمية (على ســبيل المثال، سيســكو، وآي بي أم، ومايكروســوفت، وإنتل، وســيمنز، وأوراكل، وســاب) والتي اعتمــدت بحماس من قبل المؤسســات البلدية والوطنيــة والعابرة للحدود التي تتوقع لتقنيات المــدن الذكية أن تُنتج تقــدم اجتماعي-اقتصادي وتجدد المراكز المدنية لتصبح مراكز للابتكار والعمل (Kourtit et al. 2012). وفي حين تمّ بناء بعض مشــاريع المدن الذكية من الألف إلى الياء (على ســبيل المثال، ســونغدو (Songdo)، أو مدينة مصدر (Masdar City))، فإن معظم هذه المشاريع تدريجية، وتتألف من إعادة تجهيز للبنية التحتية القائمة مع تقنيات رقمية وحلول للبيانات؛ لأن الوظيفة الرئيســية للبيانــات الكبيرة في كلتا الحالتين هي توفير تحليلات بشــكل آني أو لحظي لإدارة كيفية جوانب وظيفة المدينة وتنظيمها.

لقد استخدمت هذه المراقبة وآليات تحليل البيانات الآنية واللحظية لعدد من السنوات في بعــض القطاعات. على ســبيل المثال، قيام العديد من المــدن ببناء أنظمة النقل الذكية، والتي تســتخدم البيانات الناتجة عن شبكة من الكاميرات والنواقل والمتعلقة بحركة المرور حول نظام ما لتقديم تغذية راجعة لمركز تحكم مركزي، وتستخدم لمراقبة وتنظيم التدفق، وتعديل تسلســل حركة المرور، وحدود السرعة المســموح بهــا، وإدارة العقوبات المتعلقة بالمخالفات المرورية بشــكل تلقــائي (Dodge and Kitchin 2007a). وبالمثل، فقد ترصد وتراقب الشرطة مجموعة من الكاميرات وســجلات الحوادث الحية من أجل توجيه الموارد المناســبة لمواقع معينة للاســتجابة السريعــة وبكفاءة. كما يمكن جمــع البيانات المتعلقة بالظــروف البيئية من خلال شــبكة استشــعار موزعة في جميع أنحاء المدينة، على ســبيل المثال، قياس وتتبّع تلوّث الهواء، ومســتويات المياه، أو النشــاط الزلزالي. وتستخدم العديد من الحكومات المحلية نظم إدارية لتسجيل مشاركات الجمهور وتفاعله مع خدماتها. وفي جميــع الحالات تقريباً، كانت هــذه الأنظمة معزولة عن بعضها البعض وتتعامل مع قضية واحدة ويتم التحكم بها من قبل وكالة واحدة.

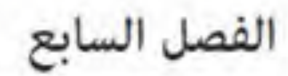
وفي الآونــة الأخيرة، كانت هناك محاولة لجمع كل هذه الأنواع من المراقبة والتحليل في مركز واحد، وأن تستكمل بآليات تحليل عامة ومفتوحة على نطاق أوسع. على سبيل المثال،



قــام مركــز العمليات في ريودي جانــيرو، البرازيل، والذي تم إنشــاءه بالشراكة بين حكومة المدينة وشركة آي بي أم، بتجهيز نظام على مستوى كافة أنحاء المدينة حيث يجمع البيانات من ثلاثين وكالة معاً في مركز موحّد لتحليلات البيانات، بما في ذلك حركة المرور والنقل العام، والبلديــة وخدمات المرافق، وخدمات الطوارئ، ونقــاط جمع بيانات الطقس، والمعلومات المرســلة من قبل الموظفــين والجمهور عبر الهاتف، والإنترنت، والراديو (انظر الشــكل 7.2، Singer 2012c). هنــا، تعمــل الخوارزميــات وفريق من المحللين عــلى معالجة، وتصوّر، وتحليل، ورصد كمية هائلة من بيانات الخدمة الحية، جنباً إلى جنب مع البيانات المجمّعة مــع مرور الزمن وكميات ضخمة من البيانات الإدارية التي تم إصدارها على أســس دورية بصورة أكبر، وغالباً ما يتم المزج بين مجموعات البيانات معا للتحقيق في جوانب معينة من حيــاة المدينة والتغييرات الحاصلة فيها مع مرور الوقت، وبناء النماذج التنبئية فيما يتعلق بالتنميــة والإدارة اليومية للمدينة وإدارة حالات الكــوارث كالفيضانات. ويكمل هذا عن طريق منصة العمليات الافتراضية التي تمكّن المســؤولين في المدينة من تسجيل الدخول من الميــدان والوصول إلى المعلومات بشــكل آني أو لحظي. فعلى ســبيل المثال، يمكن للشرطة في مشــهد الحادث المروري اســتخدام المنصة لمعرفة عدد سيارات الإسعاف المرسلة للموقع وموعــد وصولها، وتحميل معلومات إضافية على الموقــع (Singer 2012c). وكان الهدف المعلن من عمدة المدينة، إدوارد بايس (Eduardo Paes)، هو إسقاط الحواجز بين الإدارات وضمّ بيانات كلّ منها إلى الآخر لمســاعدة المنظومــة بأكملها (Singer 2012c). كما يجري تطوير مراكز مماثلة في أماكن أخرى، يصاحبها مجموعة من التطبيقات للمواطنين للوصول والاستفادة من بعض مناحي البيانات.

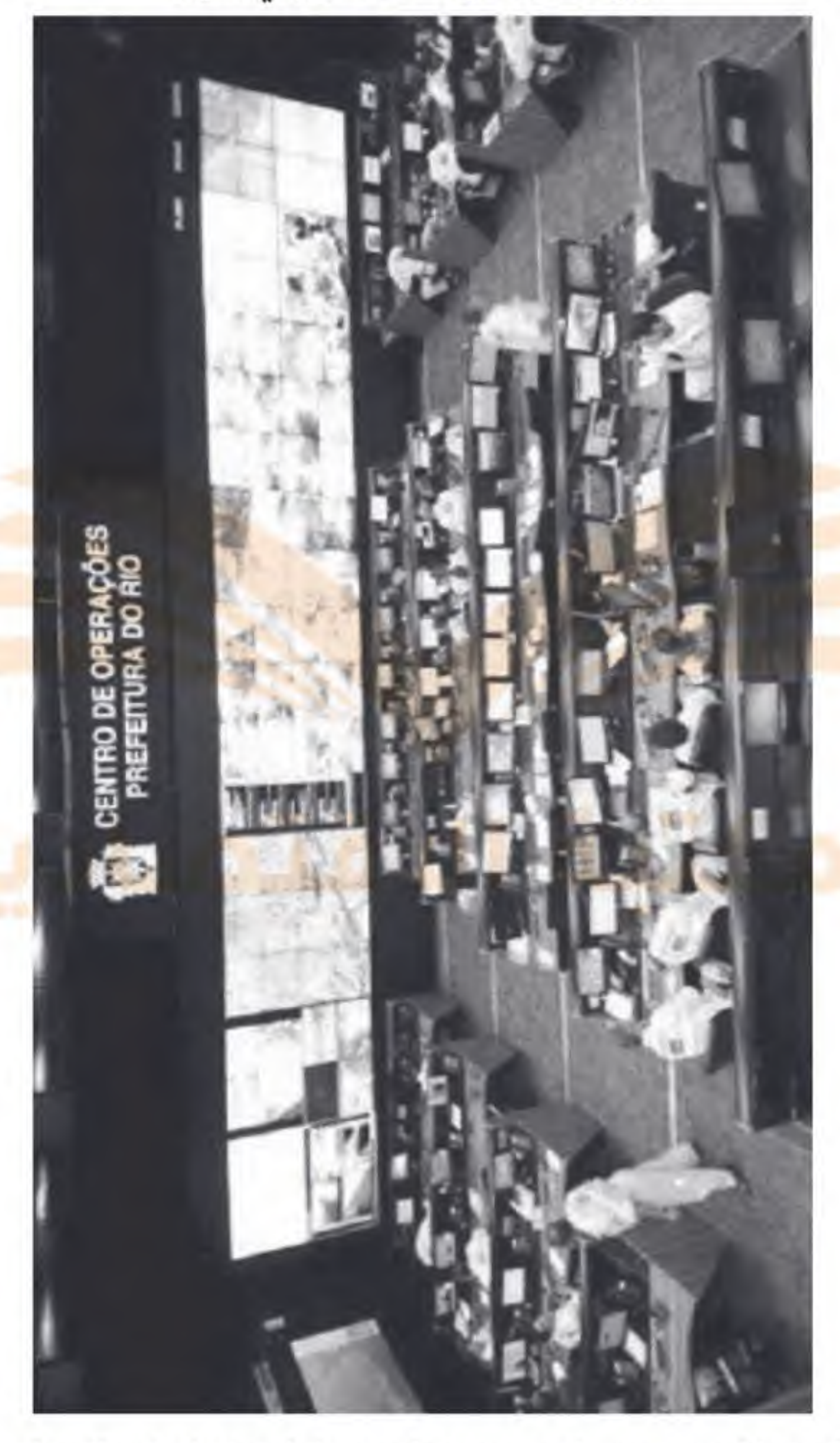
أما بالنســبة لأولئك العاملين على تطوير واســتخدام آليات تحليــل متكاملة وآنية لبيانات المدينة، فإن مثل هذه المراكز والتطبيقات توفّر وســائل فعالة لاستشعار وفهم العيش في المدينة وإدارتها في الوقت الراهن، وتصوّر وتوقّع الســيناريوهات المستقبلية، وإجــراء توزيع المــوارد بالاعتماد على الدليل وتســهيل التفكير المترابــط عبر الحكومة (Flowers 2013; Kitchin 2014). وبــدلاً مــن اتخــاذ القرارات اســتنادًا إلى الكلام المتداول، أو الحدس والظنّ، أو المحسوبية، أو أدلة دورية / جزئية، فمن الممكن تقييم ما يحدث في أيّ وقت، والرد والتخطيط المناسب له.





الشكل رقم (٢ - ٧)

مركز العمليات لوسط مدينة ريودي جانيرو



المصدر : http://ipprio.rio.rj.gov.br/centro-de-operacoes-rio-usa-mapas-feitos-pelo-ipp/

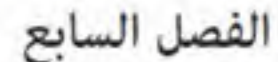


وعلاوة على ذلك، يوفّر استخدام عينات كبيرة وربط أشكال متنوعة من البيانات تحليلا أكثر عمقا، وشمولية، وقوة. وقد أصبح من الممكن للدعاة لمثل هذه الأنظمة تطوير، وتشغيل، وتنظيم، وخوض غمار الحياة في المدينة على أساس قوي، ودليل منطقي غير ضعيف، لا دليل انتقائي مبني على أسس إيديولوجية سياسية (Flowers 2013). وهكذا، فثمة من يقول، إن استخدام هذه البيانات الكبيرة سيوفر الأساس لمدينة أكثر فعالية، واستدامة، وتنافسية، وإنتاجية، وانفتاحا، وشفافية. إلا أن آخرين يحذرون من أن البيانات الكبيرة تشير إلى احتمالات الوصاية "Big Brother "، وأشكال الحكم القائم على التقنية (Technocratic)، وخصخصة (Corporatisation) إدارة المدينة (انظر الفصل العاشر).

الخلاصة:

رسم هذا الفصل الدوافع الدعائية للبيانات الكبيرة فيما يتعلق بأربع مهام رئيسية هي: حكم الشعب، وإدارة المنظمات، وتعزيز القيمة وإنتاج رأس المال، وإيجاد أماكن أفضل للحياة (حيث سيجري مناقشة إمكاناتها بالنسبة للعلوم في الفصل التالي). وقد بيّنت المناقشة الكيفية التي يتم بها تأطير البيانات الكبيرة سياسيًا واقتصاديًا (انظر الفصل الأول) وكيف يتم وضع مجموعة قوية من البيّنات والحجج العقلانية لدعم البدء في تنفيذ واعتماد تقنيات البيانات الكبيرة وحلولها. وتتضح قوة الأنظمة الخطابية التي يجري بناؤها من خلال النظر في الحجج المضادة - حيث من الصعب أن نقول إننا نريد أن نبقى في حالة أقل تبصّرًا وحكمة، وإنتاجية، وتنافسية، وكفاءة، وفعالية، واستدامة، وأمنًا... إلخ. فإذا كانت البيانات الكبيرة تقدم كل هذه الفوائد، فإن دعاة البيانات الكبيرة يجدون من غير المنطقي عدم المضيّ قدمًا في تطوير نظم البيانات الكبيرة ودعمها. وبطبيعة الحال، فإن الحجج التي قدّمت هي ضيقة وانتقائية وتبتعد عمداً عن تسليط الضوء على الآثار السلبية المحتملة فيما يتعلّق بالحريات المدنية، ومراقبة البيانات، والفرز الاجتماعي، وأمن البيانات، وانفلات زمام التحكم وفرض السيطرة، والحكم الاستباقي، والحكم التكنوقراطي (القائم على التقنية) والخصخصة، وغلق التقنيات وعدم إتاحتها





بصـورة أكبر (انظر الفصل العاشر). فهذه هي وجهـة نظر أصحاب المصالح الخاصة، ولا سـيما أولئك الذين يسـعون لبيع تقنيات البيانات الكبيرة، والحكومات التي تسعى نحو رؤيـة الليبرالية الجديدة للحكم والتنظيم (الليبرالية الجديدة: هي فكر أيديولوجي يمثل تأييد الرأسـمالية المطلقة وعدم تدخل الدولة في الاقتصاد.)، وليست وجهة نظر المواطنين أو المجتمعـات التـي من الممكن أنها لا تزال من دعاة البيانات الكبيرة والحوسـبة في كل مكان، ولكنهم يتصورون أن اسـتخدامها سـيكون بطرق تحررية، وتمكينيه، وتشاركية مع إخضاع التأثيرات الأكثر سلبية للرقابة والتنظيم.

لقد تميّز التحليل المعروض في هذا الفصل بالإيجاز، مع توفير لمحة شاملة لبعض الحجج التي تطرح. وما هو مطلوب في الوقت الحاضر، من خلال دراسـات الحالة المحددة، هو تفصيل وتمحيص للأنظمة الخطابية التي تتكشـف ويجري بناؤها. على سبيل المثال، فيما يتعلـق بالترويج للبيانات الكبـيرة في المدن الكبيرة، فهناك حاجة لرسـم وتجزئة الكيفية التـي تروّج بها مختلف الخطابات الداعمة فيما يتعلق بأماكن محددة من خلال المجلات التجارية والمواقـع الإلكترونية، والصحف اليومية، والمواد الترويجية للشركات، وسـجلات الدولة، والبيانات السياسية... إلخ، وتقاطعات الأجندات المختلفة ومصالح المعنيّين. وهذا يحتـاج إضافة إلى تحليل الكيفية التي يتمّ بها التصدي لمثل هذه الخطابات برؤى بديلة، وتحليـل المُعارضات والتداخل بين الحجج الخطابية وتتبّع تسلسـل الكيفية التي تحولت بها مع مرور الوقت، والتحليل المقارن للمشـهد الخطابي عبر مختلف البقاع، أن يصاحبه تحليلٌ لآثار الخطابات على تشـكيل الكيفية التي يعمل بها المجال في الممارسـة والنتائج المترتبة على تنفيذ حل البيانات الكبيرة.

وبالنسبة فيما يتعلق بالنقطة الأخيرة، بيّن Salmon (2014) بالتفصيل أنه بمجرد أن يتم تطبيق حل قائم على البيانات فإنه من النادر أن يحقق الفوائد التي توقعها المتحمسـون والداعمـون لذلك الحل، وغالبًا ما ينتج عن الحـل مجموعة من القضايا الجديدة. وقال إنـه توجّه لقانون كامبل (Campbell's Law)، الذي صاغـه عالم الاجتماع دونالد كامبل (Donald T. Campbell 1976)، ليـشرح هذه الظاهرة. فقانون كامبل ينص على: "إنه كلما كثُر اسـتخدام أي مؤشر اجتماعي كمي لصنع القـرار الاجتماعي، كلما كان الموضوع



تحت ضغوط الفساد وأكثر عرضة للتشويه وإفساد العمليات الاجتماعية التي يهدف إلى رصدها." أو بحسب صياغة سلمون نفسه: "كلما جرى تشغيل حقل بنظام ما بشكل أكبر، أوجد هذا النظام المزيد من الدوافع للجميع (الموظفين، والعملاء، والمنافسين) لتغيير سلوكهم بطرق منحرفة ضارة - ممّا يوفر أكثر من الشيء الذي تم تصميم النظام لقياسه وإنتاجه، سواء كان لذلك أي قيمة في الواقع أم لا".

وبعبارة أخرى، يبدأ الناس في اللعب على النظام بعقلانية واعتبار المصلحة الذاتية، ولكن في كثير من الأحيان بطرق غير متوقعة لإفساد المقاييس، والخوارزميات، وعمليات صنع القرار الآلي. على سبيل المثال، أشار Porter (1995: 44)، إلى أنه "عندما يتمّ الحكم على مديري الأعمال من خلال حساباتهم، فإنهم يبدؤون التَّعلم على كيفية تحسين حساباتهم، ربما من خلال بعض الحيل من قبيل تأجيل الصيانة اللازمة وغيرها من التكاليف ذات المدى الطويل". ونتيجة لذلك، فإنه في حين يُمكن أن تتحول كيفية إدارة وتشغيل المنظمات والحكومة، لتصبح قائمة على البيانات ومقاييس الأداء، فإنه من المحتمل أن يُحدث ذلك آثارًا متنوعة وغير متوقعة ويُفعل أشكالاً معينة من الحاكمية، ليس بالضرورة أن تكون إيجابية في مجملها.

ولذا، تحتاج دراسات البيانات الكبيرة وآثارها إلى الانفتاح لعرض السياسات والأجندات الكامنة للبيانات الكبيرة في مختلف السياقات وتوفير أساس لمزيد من المحادثات المعيارية حول هذا النوع من عوالم البيانات الكبيرة التي قد نرغب في العيش بها. فهذه الأنواع من المحادثات هي قيد التطوير في الوقت الحاضر، ويرجع ذلك بشكل جزئي إلى حداثة التطورات وسرعة وقوعها، وأيضًا بسبب هيمنة الأطروحات المبالغ بها عن البيانات الكبيرة ومكائد أصحاب المصلحة المعنيين بها. ومع ذلك، تضل الحاجة قائمة لمثل هذه الأحاديث بالنظر إلى أن البيانات الكبيرة تعيد تشكيل الطريقة التي يجري بها حكم المواطنين، وتنظيم المواقع، وإدارة المنظمات، وعمل الاقتصاديات، وممارسة العلوم.







الفصل الثامن

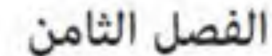
إعادة صياغة بحوث العلوم، والعلوم الاجتماعية، والعلوم الإنسانية

The Reframing of Science, Social Science and)
(Humanities Research

في حديــث لمجلة الإيكونيميســت في العــام 2010 م، ذكر ســنان آرال (Sinan Aral)، أســتاذ إدارة الأعمال في جامعة نيويورك، أن "الثورات في العلوم غالبًا ما يســبقها ثورات في القيــاس" (مقتبس عن Cukier 2010). إن هذا الفصل يستكشــف مدى انطباق ملاحظة آرال فيما يتعلق بالبيانات الكبــيرة، والبنية التحتية للبيانات، والبيانات المفتوحة، وتأثيراتها فيما يتعلق بكيفية مقاربة البحوث والاكتشــافات الأكاديمية من منظور فلســفي ومنهجي. وعلى وجه الخصوص، يقوم هذا الفصل بدراسة كيفية تحدى توافر البيانات الكبيرة والبنى التحتيــة للبيانات إلى جانب الأدوات التحليلية الجديدة للنظريات والعلوم المعرفية القائمة في مختلف التخصصات- الكيفية التي يتم بها طرح الأسئلة وكيفية الإجابة عنها- ويؤدي إلى نشــوء مجالات وتخصصات جديدة. وهكذا، يستكشف هذا الفصل ما إذا كانت التغيرات التي تطرأ على الطريقة التي يتم بها إنتاج وإدارة وتحليل المادة الخام للمعلومات والمعرفة ستؤدي إلى تغيير جذري في فهم وممارسة المبادئ الأساسية التي تقوم عليها العلوم، والعلوم الاجتماعية والإنســانية. وبعبارة أخرى، هل تقود ثورة البيانات إلى نقلة نوعية في المجالات الرئيسية للعلوم والبحوث؟

يمثّل النموذج (Paradigm)، كما حدّده كوهن (1962) Kuhn، وسيلة مقبولة لاستجواب العالم وجمع المعرفة المشتركة بين نسبة كبيرة من الباحثين في اختصاص معين في لحظة زمنية معينة. ويتشارك الباحثون العاملون ضمن نطاق هذا النموذج في فلسفة واحدة، واستخدام مجموعة من الأساليب الشائعة، وطرح الأسئلة والإجابة عنها للعمل على بناء المعرفة بصورة تدريجية. ومن ثم، فإنهم يميلون إلى تفضيل التصنيفات، والمعارف، والنظريات، والأساليب،





والأطــر الأخلاقية والعقائدية نفســها أو القريبة من بعضها البعض. ومع ذلك، فقد أشــار كوهن، وبصورة متكررة، إلى بروز طرق جديدة للتفكير تتحدى النظريات والمناهج المقبولة والمعمول بها. على سبيل المثال، فإن النظرية النسبية لآينشتاين جابهت، وإلى حد كبير، حلت محــل نظرية نيوتن في الفيزياء. وبالمثل، أحدثت نظريــة التطور لداروين تغييراً جذرياً في الفكر النظري للعلوم البيولوجية، كما تحدّت المذاهب الدينية في معتقدات النشوء والخلق. وفي كلتا الحالتين، فقد تمسّــك بعض الناس بالطرق المتبعة في التفكير وتشــبثوا بها، في حين انتقل آخرون إلى الفكر الجديد، والذي سرعان ما أصبح الفكر المهيمن. وفي بعض المجالات الأكاديمية، هناك القليل من الأدلة على نماذج عاملة، لا سيما في العديد من العلوم الاجتماعية، مثــل علم الاجتــماع والجغرافيا البشريــة، حيث تمّ تطبيق مجموعــة متنوعة من المناهج الفلســفية على ســبيل المثال، الوضعية (Positivism)، والظواهرية (Phenomenology)، والبنيويــة (Structuralism)، وما بعد البنيوية (Poststructuralism)... وغيرها، وقامت العديد من هذه المعســكرات النظرية بطرح آراء متعارضة بشأن أفضل السبل لفهم العالم. وفي مجــالات أخرى، مثل العلوم، كان الاجتماع أكثر على وحدة المنهج العلمي المتبع، معززاً باختبــار الفرضيــات للتحقق منهــا أو نفيها. وهذا لا يعني، مع ذلك، عــدم تحوّل أو تغيّر الطريقة التي تعمل على تصوير ونشر المنهج العلمي بشكل دوريّ، ووجود بعض النظريات المتنافســة فيما يتعلّق بشرح ظواهــر معينة، فالنظريات يمكــن أن تختلف حول الظواهر ولكنها تشترك في نفس النهج الأوسع للمساعي العلمية.

وقــد بــيّن جيم غراي (Jim Gray)، على ســبيل المثال (كما هو مفصّــل في .Hey et al 2009)، تطــوّر العلوم من خلال أربعة نماذج، حيث أشــار إلى أن النموذج الرابع ما زال في بداياتــه ولكــن هو نتيجة لثــورة المعلومات المنتشرة (انظر الجــدول 8.1). وبخلاف طرح كوهن، تحدث التحولات في النموذج لأن النموذج الســائد في العلم لا يمكنه تفســير بعض الظواهــر المعينة أو الإجابة عن الأســئلة الأساســية، ومن ثم تبرز المطالبــة بإعادة صياغة أفكار جديدة، فتحولات غراي تأسست على التطورات الجارية في أشكال البيانات وتطورات الأســاليب التحليلية الجديدة. وسيستكشــف هذا الفصل إلى أي مدى يأتي وجود النموذج الرابــع في العلــوم، وإلى أي مدى تقود ثورة البيانات إلى نظريــات معرفية بديلة في العلوم الإنســانية والعلوم الاجتماعية. وبالنظر إلى المراحل الأولية نسبياً في ثورة البيانات المنتشرة،

فربما لن تتفاجأ بتباين الآراء حول تأثير البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، وتحليلات البيانات الجديدة على إنتاج نظريات معرفية جديدة، وسيستمر وجود مثل هذه التباين في الآراء. ولكن، لا مجال للشك في أن البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، وتحليلات البيانات الجديدة تطرح قضايا معرفية أساسية وذلك لأن استخلاص المعلومات الصحيحة والمفيدة من طوفان البيانات لا يُعد ببساطة قضية فنية يمكن التعامل معها من خلال حلول تقنية فحسب (Floridi 2012).

الجدول رقم (١ - ٨)

نماذج العلم الأربعة

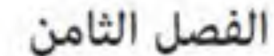
النموذج	الطبيعة	الشكل	الوقت
الأول	العلوم التجريبية	التجريبية، وصف الظواهر الطبيعية	قبل عصر النهضة
الثاني	العلوم النظرية	النمذجة والتعميم	قبل ظهور الحواسيب
الثالث	العلوم الإحصائية	محاكاة الظواهر المعقدة	قبل البيانات الكبيرة
الرابع	العلوم الاستكشافية	يتسم بكثافة البيانات، الاستكشاف الإحصائي والتنقيب عن البيانات	في الوقت الحالي

المصدر: تم التجميع من Hey et al. 2009.

النموذج الرابع للعلم (The Fourth Paradigm in Science):

في حين أن جيم غراي (Jim Gray) تصور أن النموذج الرابع في العلوم يتسم بكثافة البيانات واتساع المنهج العلمي المؤسس، إلا أن بعض العلماء الآخرين يرون أن البيانات الكبيرة أفسحت الطريق لبزوغ عصر جديد للفلسفة التجريبية (Empiricism)، حيث يمكن لحجم البيانات الذي تصحبه أساليب معينة أن يكشف عن الحقيقة الكامنة فيها، وأن يُمكن البيانات من أن تتحدث عن نفسها وهي متحررة من النظرية. وقد اكتسبت وجهة النظر الأخيرة الاعتمادية خارج المجمعات العلمية، وخاصةً داخل دوائر الأعمال، ولكن أفكارها تأصلت وتجذرت في المجال الجديد لعلوم البيانات والعلوم الأخرى. فكلا المنهجين





يوفـران تحدياً للمنهجية العلمية التقليديـة إضافةً إلى طرح نموذج بديل محتمل، وفي هذا الجزء، سيتم دراسة الادعاءات العلمية لكلا المنهجين بشكل نقدي.

عودة ظهور الفلسفة التجريبية (The Re-Emergence of Empiricism):

بالنسـبة للبعض، فإن فلسـفة النموذج الرابع هي ما أطلق عليـه (Brooks (2013 a «منهج البيانات» (Data-ism)، وهو محاولة التقاط أي شيء على شكل بيانات واستخلاص مستوى معين من التبصر منها: "فأي شيء يمكن قياسه يجب أن يتم قياسه... البيانات هي منظور شـفاف وموثوق يسـمح لنا بتنقية العاطفية والأيدلوجية؛ إن البيانات ستسـاعدنا على فعل أشـياء رائعة - مثل التنبؤ بالمستقبل". إن من شأن مثل هذه الادعاءات المتعلقة بالبيانـات أن تعزز كثيراً من الدعايات بشـأن البيانات الكبيرة داخل مجتمع الأعمال، ويتم التعبير عنها بشـكل عام مـن خلال الصياغة التجريبية - التي يمكـن لها من خلال الحجم الكافي للبيانات أن تعبر البيانات عن نفسها. ومثل هذه الفلسفة التجريبية متضمنة بشكل أمثـل في ادعـاءات (Chris Anderson (2008، وهو رئيس تحرير سـابق في مجلة وآيرد (Wired)، والـذي ضربت دعوته التأليبية بأن البيانات الكبيرة ما هي إلا مؤشر على "انتهاء النظرية" على وتر حساس للكثير من المعلقين. لقد ذكر Anderson في مقطوعة استفزازية أن "طوفـان البيانات يجعل المنهج العلمـي مهملاً"، وأن الأنماط والعلاقات المتضمنة داخل البيانـات الكبيرة تقدم بشـكل متأصـل معرفة مهمـة وثاقبة عن العمليـات الاجتماعية، والسياسية، والاقتصادية، والظواهر المعقدة. وقد أكد ذلك قائلاً:

> هناك الآن طريقة أفضل. فالبيتابايت تسمح لنا بأن نقول: "إن الارتباط كافٍ" (Correlation is enough). ويمكننا أن نتوقف عن البحث عن النماذج. ويمكننا أيضًـا أن نحلل البيانات بدون أي افتراضات بشـأن ما يمكن أن يظهره التحليل لنا. ويمكننـا أن نرمي الأرقام في أكبر تجميعات الإحصاء التي شـهدها العالم يوماً، وأن نـدع الخوارزميـات الإحصائية تجد الأنماط إذا كان العلم لا يسـتطيع إيجاد هذه الأنمـاط... فالارتباط يحل محل السـببية، ويمكن للعلم أن يتقـدم ويتطور حتى بدون النماذج المترابطة أو النظريات الموحدة، أو أي تفسـير آلي على الإطلاق. لذا، لا يوجد أي سبب للتعلق بالطرق القديمة.

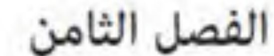


وعلى نحو مشـــابه، أكد Prensky (2009) على أنه "لم يعد يتعين على العلماء الآن القيام بتخمينات تنم عن الثقافة، أو أن يقوموا بتأسيس افتراضات ونماذج، ويختبروا هذه الافتراضات وتلك النــماذج من خلال التجارب والأمثلة المعتمدة على البيانــات. فبدلاً من ذلك، يمكنهم التنقيــب في المجموعــة الكاملة من البيانات بحثــاً عن الأنماط التي تُظهــر التأثيرات وتُقدم النتائــج العلمية بدون إجراء أي تجريب إضــافي". ومن ثم فلقد أكدت Dyche (2012) أن: "التنقيب في البيانات الكبيرة يكشــف العلاقات والأنماط التي لم نكن نعرفها كي نبحث عنها". واستمرت في الادعاء قائلةً: "هذه الأنماط محددة للغاية وتبدو وكأنه تم التحكم في تحديدها من خلال الافتراضات". وبالمثل، أكد Steadman (2013)، على ما يلي:

> يُتيح منهج البيانات الكبيرة تجاه جمع المعلومات المبني على المعرفة للمحللين الحصول على القرار الكامل فيما يتعلق بالشــئون عالمية النطاق. فلا يتم فقد أي شيء عنــد النظــر عن قرب بالغ إلى جزء واحد محدد من البيانات؛ ولا يتم فقد أي شيء عند محاولة الحصول على منظور واسع للغاية بشأن موقف حيث تم فقدان التفاصيــل الدقيقــة. فالخوارزميات تجد الأنماط في حــين أن الافتراضات تتبع من البيانات، ولا يجب على المحلل أن يزعج نفســه حتى بطرح افتراض بعد الآن. لقد تحــول دورها من دور اســتباقي إلى دور تفاعلي، حيــث إن الخوارزميات هي من تقوم بالعمل الســياقي... فالخوارزميات ســتقوم بتحديد الأنماط ثم تقوم بوضع النظريــات، ولذا فهناك حاجة متناقصة للقلق بشــأن وضع افــتراض في البداية ثم اختبار هذا الافتراض من خلال عينة من البيانات.

وبعبارة أخرى، فبدلاً من اختبار إذا ما كانت بعض الأنماط المفترضة المعينة أو العلاقات توجــد داخل مجموعة بيانات، فإن الخوارزميــات مصممة للعمل على البيانات الكبيرة من أجل اكتشــاف ارتباطات ذات دلالة بين البيانات دون أن يتم توجيهها من قبل الافتراضات. وكما قال Croll (2012:56): "في النموذج القديم المتســم بندرة البيانات، كان يتعين علينا أن نقرر ما الذي ينبغي جمعه أولاً ثم نقوم بجمعه. أما مع النموذج الجديد المتسم بوفرة البيانات، فنحن نقوم بالجمع أولاً ثم نسأل الأسئلة لاحقاً".





إن الأمثلة المستخدمة في توضيح مثل هذا الموقف عادةً ما تكون نابعة من التسويق والتجزئة. فعلى سبيل المثال، وصف Dyche (2012) حالة سلسة تجزئة حيث قام بتحليل صفقات الشراء لمدة اثني عشر عاماً من أجل ملاحظة العلاقات المحتملة غير الملحوظة بين المنتجات التي انتهى بها الأمر في سلة المشتريات للمتسوقين. وقد أدى اكتشاف الارتباطات بين البنود المعينة إلى تعيين الأماكن الملائمة للمنتج الجديد وإجراء بعض التعديلات على إدارة مساحة الأرفف مع زيادة تبلغ ١٦٪ في الإيرادات لكل حقيبة تسوق في الشهر التجريبي الأول. لم تكن هناك أي افتراضات بأن المنتج (س) على سبيل المثال غالباً ما يتم شراؤه مع المنتج (ص) والتي تم اختبارها فيما بعد. لقد تم بكل بساطة طرح بعض التساؤلات بشأن البيانات من أجل اكتشاف ماهية العلاقات الموجودة التي لم يكن من الممكن ملاحظتها في السابق. وبالمثل، فإن نظام التوصيات في موقع أمازون يقدم اقتراحات بشأن منتجات أخرى من الممكن أن يكون المتسوق مهتماً بها دون معرفة أي شيء عن ثقافة أو تقاليد الكتب والقراءة لذلك المتسوق، فهذا النظام يحدد، ببساطة، أنماط الشراء عبر العملاء، وذلك من أجل تحديد ما إذا كان الشخص (س) يحب الكتاب (ص)، فمن المحتمل أنه سوف يحب الكتاب (ز) وذلك بالنظر إلى الأنماط الاستهلاكية له وللآخرين.

ويزعم Dyche أن هذا المنهج المفتوح للاكتشاف، بدلاً من أن يكون موجهًا، يُعد أكثر احتمالية لكشف الأنماط الضمنية غير المعروفة، وذلك فيما يتعلق بسلوكيات المستهلك، والعلاقة بين المنتجات، والمخاطر المالية، والتي يمكن الاستفادة منها لاحقاً. وفي حين أن من الممكن أن يكون من المفضل توضيح أي ارتباطات موجودة داخل البيانات وما السبب في كون هذه الارتباطات ذات دلالة، فإن مثل هذا التفسير يُعد غير ضروري بشكل كبير، فالمهم هو معرفة أنه يفعل ذلك. ومن ثم فلقد ادعى Siegel (2013: 90)، (هذا هو تأكيدي) فيما يتعلق بالتحليلات التنبئية: "نحن في العادة لا نعرف عن السببية، ونحن في الغالب لا نهتم بذلك بالضرورة... إن الهدف يركز بشكل كبير على التنبؤ أكثر من تركيزه على فهم العالم... إنه بحاجة إلى أن ينجح فحسب؛ إن التنبؤ يفوق التفسير". وكما ادعى Weinberger (2011: 33) فيما يتعلق بهُنش (Hunch)، وهو موقع شبكي يقوم بتقديم توصيات بناء على الإجابات عن الأسئلة التي يبدو أنها لا ترتبط ارتباطاً وثيقاً بالأسئلة الأصلية التي تم توجيهها:

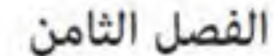


إن التحليــل لا يدعم النظريــة وهو لا يقدم أي نظريــة. لا توجد لدى موقع هُنــش (Hunch) أي فكــرة لم أن الناس الذين، على ســبيل المثال، يفضلون ارتداء صندل على الشــاطئ والذين لم يقوموا بنفح الهندباء (نبته بريه) في العام الماضي، ربمــا يحبون هذه الأفلام الأربعة (التي تم اقتراحها لهم). فالموقع لم يكن لديه أي افتراض، ولم يكن لديه أي تخمين. كل ما كان لديه ارتباط إحصائي فحسب.

إن بعــض برامــج تحليلات البيانــات يتم بيعها مــن أجل هذا التصــور بالضبط، فعلى ســبيل المثال، فإن البعض يدعي أن برنامج التنقيب عن البيانات والعرض البصري آياسداي (Ayasdi)، لديــه القدرة على اكتشــاف الرؤى بشــكل آلي بغض النظر عــن التعقيد ودون توجيه أي أســئلة. ويمكن لعملاء آياســداي أن يتعلموا أخيراً الإجابات عن الأســئلة التي لم يكونوا يعرفون أن يسألوها في المقام الأول. ولعرض الأمر بصورة مبسطة، فإن آياسداي هو «سرنديبية رقمية» بمعنى أن آياســداي لديه القدرة على اكتشــاف الأشياء المفيدة مصادفة (من أسطورة أمراء سرنديب الثلاثة)(Clark 2013). وعلاوةً على ذلك، فهناك من يدعي أن آياســداي أزال بشــكل كلي دور العنصر البشري في عملية التنقيب عن البيانات – ومن ثم، فلقد أزال جميع التحيزات البشرية المتضمنة معه. وبدلاً من انتظار أن يتم توجيه ســؤال إليه، أو أن يتم توجيهه إلى روابط بيانات موجودة معينة، فإن النظام ســوف يقوم وبدون توجيه باكتشاف الأنماط التي ربما لم يفكر المتحكم البشري أن يبحث عنها (Clark 2013).

إن هنــاك مجموعة قويــة وجذابة من الأفكار التي تعمل في مثل هذه المجادلات والتي تجري على نحو متضاد مع المنهج الاستدلالي المسيطر والمهيمن داخل العلوم الحديثة. أولاً: أنه يمكن للبيانات الكبيرة أن تلتقط المجال بالكامل وأن تقدم قراراً كاملاً. ثانياً: أنه لا توجد هنــاك حاجة للنظرية، أو النماذج، أو الافتراضات الاســتنباطية. ثالثاً: أنه من خلال تطبيق تحليــلات البيانــات بدون دراية يمكــن للبيانات أن تتحدث عن نفســها وهي متحررة من الصياغة والانحيــاز البشريين، وأن أي أنماط أو علاقات داخل البيانات الكبيرة تكون صادقة وهادفة بشكل أصلي. رابعاً: أن المعنى يتجاوز السياق أو المعرفة المرتبطة بالمجال.

فهذه الأمور تعمل ســوياً لكي تشير إلى أن هناك أسلوبًا جديدًا في العلوم يتم تأسيسه، وهو الأســلوب الذي تكــون فيه طريقة العمل اســتقرائية بطبيعتهــا، وبمثل هذا ادعى

Gannon and Reid (2009) أن "التوافــر الجاهــز في البيانــات المتعددة يحول المناهج العلمية من الطريقة العلمية التقليدية التي تقودها الافتراضات إلى العلوم المعتمدة على الاستكشــاف". وبالمثل، يقــترح (127 :2011) Weinberger أن الهدف التقليدي للعلوم المتمثل في تكوين النظريات التي تدعمها الحقائق وتفسر الحقائق في الوقت ذاته يتحول في الوقــت الراهــن إلى "تقديم نمــاذج معقدة تفسر العلاقات المعقــدة" حيث "في بعض الحالات يكون من المســتحيل كشــف الســبب وراء عمل تأثيرات معينة، وبدلاً من ذلك، هنــاك ارتبــاط قوي ملاحظ بين المحاكاة والعالم الحقيقي والــذي يمنح الثقة بأن المحاكاة الأخرى ســوف تقدم نتائــج موثوقة". وقد ســتخدم Weinberger المثال المتعلق بوضع عمود على بعد متر واحد أمام أحد الأبواب في برنامج محاكاة لتدفق خروج الناس في حال الإخلاء والذي أدى إلى تدفق أفضل، ولكن دون وجود أي تفســير من النموذج للسبب الذي يجعل العمود يفعل ذلك. واقترح، بطرق أخرى متعددة، أننا لسنا بحاجة لأن نعرف السبب، وإنما نحتاج لأن نعرف أن العمود يفعل ذلك وأن تطبيق الفكرة ســوف يســاعد على إنقاذ حياة الناس، وقد أنتج هذا لـ (26 :2013) Andrejevic حالة تسمى "المعرفة بدون الفهم".

مغالطات الفلسفة التجريبية (The Fallacies of Empiricism):

لا يوجد هناك أي شــك بأن البيانات الكبيرة لديها، وســوف يكون لديها بشكل متزايد، تأثير على كيفية ممارسة وفهم العلوم. وعلى الرغم من ذلك، فإن رؤية الفلسفة التجريبية بشــأن كيفية تكون هذا العلم الجديد تعتمد عــلى التفكير المغالط - جميع الأفكار الأربع التي تعزز تشــكيلها مشــكوك فيها. فلربمــا تكون هذه الأفكار جاذبة بشــكل كبير، ولكن التدقيق النقدي والفلســفي يكشــف مشــاكل خطيرة مع كل فرضية. إن من المهم أن يتم التحقــق من كل فكرة، كل بدروهــا، قبل صياغة رؤية بديلة للعلم الــذي تقوده البيانات والذي يمزج بين سمات الإبعاد، والاستقراء، والاستنتاج.

أولاً: يمكــن للبيانــات الكبيرة أن تلتقــط المجال بالكامل وأن تقدم قــراراً كاملاً. حيث إن البيانــات الكبيرة تكافح من أجل الشــمولية، فمناصروها يدعون أنه من الممكن رؤية كل شيء داخل المجال (على سبيل المثال جميع الأنشطة على موقع تويتر، أو جميع عمليات الشراء داخل سلاســل المتاجر الكبيرة)، وأنه من الممكن أن نفهم بشــكل كامل ما نــراه (Vis 2013). وعلى

الرغم من ذلك، وكما ناقشنا في الفصول ١ و٢ و٩، فإن البيانات الكبيرة ربما تسعى إلى أن تكون شمولية، ولكنها تشكل كلاً من التمثيل والعينة. فالواقع، أن جميع البيانات تقدم رؤى بصرية قليلة للعالم، وليس رؤى شاملة: أي رؤى من وجهات نظر معينة- باستخدام أدوات معينة، وذلك بدلاً من رؤى بصرية شاملة لا تخطئ (Haraway 1991; Amin and Thrift 2002). وعلاوةً على ذلك، فإن المجالات تتطور وتتغير، ومن ثم فإن البيانات الكبيرة ليس بإمكانها الوصول إلى كل مكان، وذلك يرجع من ناحية إلى أن البيانات الكبيرة لا يمكنها الإلمام بجميع الأماكن والأزمنة، ومن ناحية أخرى؛ فإن الكثير من الأماكن والأزمنة الجديدة تبقى لكي يتم اختراعها (Amin and Thrift 2002: 128). وعلى هذا النحو، تشكل البيانات الكبيرة "سلسلة من الأوامر الجزئية والكليات المتمركزة، مع مقدرتها على التحديق في بعض الاتجاهات دوناً عن الأخرى" (Latour والمقتبس في Amin and Thrift 2002: 92)، ومن ثم فلقد استنتج (Vis 2013): "أنه من المهم أن نتذكر أن ما نراه تتم صياغته من قبل ما نحن قادرون على رؤيته أو في الواقع ما نريد رؤيته من خلال إطار عمل فكري محدد".

ثانياً: أنه لا توجد هناك حاجة للنظرية، أو النماذج، أو الافتراضات الاستنباطية. الافتراض هنا هو أن تكوين البيانات وتحليلات البيانات واللذان يعززان استخدام البيانات الكبيرة يبدو أنهما يأتيان من لا شيء، وأنهما بشكل ما متحرران من "القوة التنظيمية للفلسفة" (Berry 2011: 8). فالأنظمة المتناقضة تُصمم من أجل التقاط أنواع معينة من البيانات، وغالباً ما يحصل ذلك من أجل أغراض محددة للغاية، وأن التحليلات والخوارزميات اللتين تدعمان البيانات الكبيرة تعتمدان على الاستدلال العلمي وأنه تم صقلها من خلال الاختبار العلمي. فالعديد من الأفكار المستخدمة في تصميم نظم البيانات الكبيرة مستمدة من الأدبيات العلمية واسعة النطاق والتجارب والمعارف لهؤلاء الذين يعملون على النظام. إن هناك الكثير من الوقت المستنفد في اختيار واختبار طرق متعددة لتحديد واستخلاص القيمة من البيانات المنتَجة، وتقييم النتائج وتطويع التحليلات. وهناك جدول أعمال يقود مثل هذا العمل وذلك فيما يتعلق بماهية النتائج المرجوة (زيادة دوران رأس المال أو الأرباح، أو فهم نظام معين، إلخ)، والذي يستفيد من النظريات المبرهنة، وذلك فيما يتعلق بماهية الطرق الملحوظة لتحقيق مثل هذه التأثيرات، وما إذا كانت مثل هذه التأثيرات تحدث في الواقع ثم يتم اختبارها بعد ذلك.

وكما أشـار (2012) Leonelli، فإن الإسـتراتيجية الاسـتقرائية لتحديـد الأنماط داخل البيانـات عادةً ما يقودها الاسـتدلال العلمـي، وهي لا تحدث فراغًا علميًـا. أو، كما ذكر (2012) Quintero، "لا يمكن تقديم التفسـيرات ببساطة من التحليلات الإحصائية، وذلك دون الحاجة إلى نظرية سـابقة من نوع ما". وقد أتبع ذلك قائلاً: "كونك منقاداً بالبيانات يعد أمراً مفيداً فحسب إذا كان لديك نظرية قوية يمكنك من خلالها أن تستمر في طريقك، إن أي شيء آخر سوف يتركك متجهاً إلى منحدر دون أن تراه". فالواقع، أن كلاً من الاستدلال الاسـتنتاجي والاسـتقرائي يتم تشـكيلهما دوماً بشكل استطرادي ولا ينشـأ أحدهما من لا شيء. ومـن ثم فقـد اقـترح (1979) Popper، مقتبس مـن (2012: 74) Callebaut أن جميـع العلوم تطبق منهج الأضواء الكاشـفة على الاكتشـاف العلمـي، حيث يكون تركيز الضـوء منقاداً من خلال النتائج السـابقة والنظريات والتدريب، وكذلك من خلال التخمين المؤسـس على التجارب والمعارف. إن الأمر ذاته ينطبق عـلى أمازون (Amazon) وهُنش (Hunch)، وآياسـداي (Ayasdi)، وجوجـل (Google). فالكيفية التي اعتمد عليها أمازن في بنـاء نظام التوصيات على الاسـتدلال المنطقي، كانت مُعززة من خلال نموذج إرشـادي يرافقـه اختبار تجريبي تم تصميمه من أجل تحسـين أداء الخوارزميات التي يسـتخدمها. وبالمثل، فإن جوجل يقوم ببحوث وتطوير واسـع النطاق، وهو يعمل بالشراكة مع العلماء ويشـتري المعرفة العلمية، سـواء تم هذا من خلال تمويل البحوث داخل الجامعات أو من خـلال شراء المعرفات الإلكترونية للشركات الأخرى، وذلك من أجل تنقية وتوسـيع المنفعة بشأن كيفية تنظيمه، وتقديمه، واستخلاصه للقيمة من البيانات. ونتيجة لذلك، فإذا وجدت الخوارزميات الإحصائية أنماطاً في البيانات، فإن ذلك يعود إلى أنه تم استخدام علم التعرف على الأنماط، جنباً إلى جنب مع المعرفة الخاصة بالمجال.

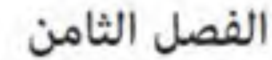
ثالثاً: يمكن للبيانات أن تتحدث عن نفسها وهي متحررة من الصياغة والانحياز البشريين. إن المفهوم القائل بأن إنتاج المعرفة من البيانات الكبيرة يحدث بشـكل غير مرتبط بالعلم، هو على علاقة بالفكرة القائلة بأن تحليلات البيانات الكبيرة تُمكن البيانات من التحدث عن نفسـها، وهي غير مثقلة بالسـياقية أو أهواء التوضيح البشري. فليس فقط إنتاج البيانات هو ما يتم بشـكل افتراضي ومتحرر من النظرية، فحتى تفسـير ومعاني تلك البيانات يمكن بطريقة مشابهة أن يحدث في فراغ علمي، فمثل هذا المفهوم يُعد سمة من سمات التفكير



التجريبي، وهو أنه عندما يتم تقديم وتحليل البيانات بطريقة ملائمة، فإن البيانات، بنفسها ومن نفسها، تشكل دليلاً ملحاً يرتبط بظاهرة أو عملية معينة، بدلاً من تفسيرها. وفي حالة البيانـات الكبيرة، فإن حجم البيانات يضيـف إلى أهمية الدليل. فهذا "التقديس" للبيانات يمكـن أن يـؤدي، كما ادعى Jenkins (2013)، إلى حتمية البيانات في المناقشـات التي تتم مـن أجل البيانات الكبيرة. وهذا يعنـي أن البيانات تتفوق على النظرية، وأن البيانات أمر حقيقي، في حين أن النظرية ليست حقيقية (Hales 2013)، وأن البيانات تقول الحقيقة في حـين أن النظرية مجرد تلفيق. وهنا، يوجـد ثلاث قضايا يتعين التفكير فيها: القضية الأولى هـي افتراض أن البيانات موضوعية، ومحايدة، ومتحررة من التحيز البشري. القضية الثانية أن الأنماط والعلاقات داخل البيانات الكبيرة هادفة وصادقة بشكل متأصل. القضية الثالثة أن تفسير مثل تلك الأنماط والعلاقات يفتقر إلى أي نوع من الوضعية والموقعية.

وكما ناقشـنا في الفصل الأول، فإن البيانات ليست ببسـاطة عناصر محايدة أو طبيعية بحيـث يتم فصلها عن العـالم بطرق محايدة وموضوعية، ويمكن قبولها بالقيمة الاسـمية. فالبيانات لا توجد قبل تكوينها ولا تنشأ من أي شيء. ولكن الأحرى، أن البيانات يتم إنشاؤها داخل مجاميع البيانات المعقدة التي تشكل تركيبها بشكل فعال، ولا يمكن للبيانات في هذه الحالة التحدث عن نفسـها، ولكنها على الدوام تتحدث، بشكل متأصل، من وضعية معينة (Crawford 2013). وعلاوةً على ذلك، فإن ادعاء Anderson (2008) بأن الارتباط يفوق السـببية يشير إلى أن الأنماط الموجودة داخل مجموعة البيانات ذات معنى بشكل متأصل. وهـذا افتراض يعرف جميع الإحصائيـون المدربون أنه خطر وغير صحيح. فالارتباطات بين المتغيرات داخل مجموعة البيانات يمكن أن تكون عشـوائية بطبيعتها ويوجد بينها ارتباط عرضي قليل، أو لا يوجد هذا الارتباط على الإطلاق (انظر الفصل التاسـع). إن تفسـير أي ارتبـاط عـلى أنه ذو مغزى ربما يؤدي في هذه الحالة إلى مغالطات بيئية خطيرة. ويمكن أن يتفاقـم هذا الأمر في حالة البيانات الكبيرة، وذلك يرجـع إلى كون الموقف التجريبي يظهر وكأنه يعزز ممارسة تجريف البيانات – البحث عن أي ارتباط – ومن ثم يزيد من احتمالية اكتشـاف ارتباطات عشـوائية. وعلى الرغـم من ذلك، ففي حـين أن التعرف على الأنماط ربما يحدد علاقات شـيقة محتملة، إلا أن صحة هذا الأمر تسـتلزم اختباره على مجموعات بيانـات أخـرى وذلك من أجل التأكد مـن مصداقيتها وصحتها. وبعبـارة أخرى، يتعين أن





تشكل العلاقات أساس الافتراضات التي يتم اختبارها على نطاق أوسع، والتي تستخدم بعد ذلك من أجل بناء وصقل النظرية التي تفسرها، ومن ثم فإن الارتباطات لا تفوق السببية، ولكنها بالأحرى يجب أن تشكل أساس بحوث إضافية من أجل معرفة ما إذا كانت مثل هذه الارتباطات دالة على السببية، وفي هذه الحالة فقط يمكننا أن ندرك مدى مغزى أسباب الارتباط.

وفي حين أن الفكرة القائلة بأن البيانات تتحدث عن نفسها وهي متحررة من التحيزات أو أن الصياغة ربما تبدو وكأنها فكرة جاذبة، إلا أن الواقع مختلف إلى حد ما. فبحسب ما أشار Gould (1981: 166)، فإن "البيانات غير الحية لا يمكنها أبداً أن تتحدث عن نفسها، ونحن دوماً نحاول أن نوجه بعض إطار العمل المفاهيمي، سواءٌ كان بديهياً ومشكلاً بصورة سيئة، أم مركبًا بشكل محكم ومنهجي، نحو مهمة الاستقصاء، والتحليل، والتفسير". إن فهم البيانات دوماً مؤطر؛ ويتم فحصها من خلال منظور معين يصوغ كيف تتم ترجمتها، وحتى لو كانت العملية مؤتمتة بطريقة ما، فإن الخوارزميات المستخدمة لمعالجة البيانات عادةً ما تكون متشبعة بقيم معينة وتكون مقترنة بالسياق داخل منهج علمي معين. وكما أوضح Hales (2013): فإن "أي اختبار إحصائي أو خوارزمية تعلم آلي تعبر عن منظور ماهية النمط أو النظام... فأحد الخوارزميات سوف تجد نمطاً ما، في حين تجد خوارزمية أخرى شيئاً آخر". فإذا ترجم ذاك شخص ما، فسيكون من المحتم أن يستفيد من معرفته، ومهاراته، وحدسه المتراكم من أجل تنقية رؤاه، حتى يقوم محلل مختلف بطرح نتائج مختلفة من نفس التحليلات. فالفهم أو التفسير يتشكل في العادة من خلال أعين المشاهدين بغض النظر عن ادعائهم أنهم محايدون أو متحررون من القيمة الناتجة، وعلاوةً على ذلك، فإن التفسيرات ليست ثابتة، فهي تتغير كلما تعرضت إلى معارف جديدة من خلال بحوث، أو مناقشات، أو مجادلات إضافية. وعلى هذا، فإن البيانات لا تتحدث عن نفسها أبداً، وكما ادعى Silver (2012: 9): "نحن نتحدث بالنيابة عن البيانات".

رابعاً: المعنى يتجاوز السياق أو المعرفة المرتبطة بالمجال. إن من الجوانب المرتبطة بالفكرة القائلة بأن البيانات تتحدث عن نفسها المفهوم القائل بأن تفسير البيانات الكبيرة لا يتطلب معرفة سياقية أو مرتبطة بمجال معين. وبدلاً من ذلك، فالمهم هو المقدرة على



توضيح المعنى داخل البيانات، فبسبب أن البيانات من المفترض أن تتحدث عن نفسها، يمكن لأي شخص ذو فهم معقول بالإحصاء أن يكون قادراً على تفسيرها. إن هناك شكوكاً قليلة فيما يتعلق بأن علماء البيانات والحاسب والعلماء الآخرين يتحركون داخل نطاق تخصصات بعضهم البعض مع انتشار البنى التحتية للبيانات والحوسبة. وفي الكثير من الحالات، فإنهم لا يقدمون أدوات جديدة فحسب، ولكنهم يشتركون بفعالية في تفسير البيانات. ومع ذلك، فإن هؤلاء العلماء لديهم القليل من الفهم المعمق في المجال المناط بهم وكذلك بالنظريات التجريبية المتعمقة والمعارف التي تم تأسيسها عبر فترات زمنية طويلة. فعلى سبيل المثال، فيما يتعلق بالعلوم الاجتماعية والإنسانية، فإن الأساليب الجديدة من التحليلات والبنى التحتية للبيانات توجه من قبل علماء الحاسب والفيزياء. وفي الغالب تجعل النتائج أولئك المتشربين لوجهة النظر التخصصية يستسلمون وينقادون لهذه النتائج.

ولتوضيح هذا الجانب، فإن المجال الناشئ للفيزياء الاجتماعية، والذي من خلاله يقوم الفيزيائيون وآخرون باتخاذ قرارات بشأن العمليات الاجتماعية والمكانية بناءً على تحليلات البيانات الكبيرة، وخاصةً تلك المتعلقة بالمدن والقوانين المفترضة التي تعزز صياغتها ووظائفها (Bettencourt et al. 2007;Lehrer 2010; Lohr 2013)، غالباً ما يتجاهل عن عمد ثقافة علم الاجتماع الممتد لقرنين من الزمان، ويشمل ذلك تجاهل التحليلات الكمية والنماذج المبنية لما يقارب قرناً من الزمان. والنتيجة هي تحليل المدن الذي يكون اختزالياً إلى حد كبير، ونفعياً ويتجاهل تأثيرات الثقافة، والسياسة، والحوكمة، ورأس المال، وتقاليد ثرية من العمل الذي سعى نحو فهم كيفية عمل المدن من النواحي الاجتماعية، والثقافية، والسياسية، والاقتصادية والنتيجة إعادة إنتاج الأنواع ذاتها من أوجه القصور التي أوجدتها العلوم الاجتماعية الكمية / الوضعية في منتصف القرن العشرين (Kitchin 2914; Mattern 2013). وهذا أيضًا يتعدى حدود التقاليد المتنوعة والثرية للإرث الفلسفي ويعمل على تعزيز المناهج الفلسفية التجريبية والوضعية بصفتها أفضل وسيلة لفهم المدن مع أنه من المحتمل ألا يتفق مع هذه الرؤية النظرية إلا القليل جداً من العلوم الاجتماعية التي تلت عصر الفلسفة الوضعية (Kitchin 2013).

إن هناك مجموعة متشـابهة مـن المخاوف لدى أولئك العاملـين في مجال العلوم. فقـد أشـار Strasser (2012) على سـبيل المثال، إلى أنه في العلـوم البيولوجية، ادعى متخصصـو المعلوماتية الحيوية الذين لديهم أسـلوب معين وضيـق للغاية لفهم علم الأحياء، أن الأرض كان يسكنها في السابق الأطباء السريريون وعلماء الأحياء التجريبيون والجزيئيون. وأشار إلى أن: "تحويل التحليل إلى البيانات الرقمية... يُفسح المجال أمام الأسئلة المعرفية كتلك المتعلقة بمن يملك الشرعية ليكون منتجاً للمعرفة- المهتم بالآثار (أو الطبيـب السريـري أو عالم البيولوجيا الجزيئية) أو الإحصـائي الذي يحلل البيانات ويقدم المحاكاة أو النموذج (2012: 7). ومن ثم فإن بعض علماء البيانات بدون شـك يتجاهلون ملاحظات Porway (2013)، وهي:

> بدون توافر خبراء في موضوع البحث من أجل توضيح المشـاكل مسبقاً، فسوف نحصل على نتائج ضعيفة... وبدون شـك فإن هنـاك حاجة إلى خبراء في موضوع البحث من أجل تقييم نتائج العمل وخاصةً عندما يتعامل الفرد مع بيانات حساسة عن السـلوك البشري. وبوصفنا علماء بيانات، فنحن مؤهلون بشكل جيد للتعامل مع "ماذا" فيما يتعلق بالبيانات، ولكن لا يجب علينا إلا فيما ندر التعامل مع سؤال "لماذا" فيما يتعلق بالموضوعات التي لا نكون خبراء فيها.

وكما أشـار Porway، فـإن المطلوب بالفعل هو أن يعمل علـماء البيانات وخبراء النطاق مع بعضهم البعض من أجل التأكد من أن تحليلات البيانات المستخدمة مفهومة وأن النتائج من مثل هذه التحليلات يتم تفسـيرها بعقلانية وسـياقية. وبالمثل، يدعو Lazer et al. (2009: 10 – 11) إلى التعاون بين علماء الاجتماع الضالعين إحصائياً وبين علماء الحاسـب الضالعين حاسـوبياً، (2009: 10 – 11)، وفيما يتعلق بالأعمال، ادعى Minelli et al. (2013) أن فـرق علـماء البيانات ينبغي أن يعملوا مع خبراء عمليات الأعمال من أجل الاستفادة من الرؤى المناسبة (انظر أيضًا الجدول ١-٩).

العلم المُقاد بالبيانات (Data-driven Science):

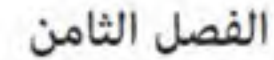
يسعى العلم المقاد بالبيانات، بدلاً من كونه متجذراً في الفلسفة التجريبية، إلى التمسك بمعتقـدات الطريقـة العلميـة، ولكنه أكثر انفتاحاً لاسـتخدام مزيج مختلـط من المناهج



الاستدلالية، والاستقرائية، والاستنتاجية، من أجل تطوير فهم الظاهرة. وهذا يختلف عن التصميم الاستدلالي التجريبي التقليدي في أنه يسعى إلى طرح افتراضات ورؤى يتم توليدها من البيانات بدلاً من توليدها من النظريات (Kelling et al. 2009: 613). وبعبارة أخرى، فهو يسعى إلى دمج أسلوب من الاستقراء في تصميم البحوث، وذلك على الرغم من أن الشرح من خلال الاستقراء ليس هو الغاية المستهدفة (كما هي الحال مع المناهج التجريبية). وبدلاً من ذلك، فهو يشكل أسلوباً جديداً من أساليب توليد الفرضيات قبل استخدام المنهج الاستنتاجي. ولا تنشأ عملية الاستقراء من لا شيء، ولكنها تقع وتقترن بسياق نطاق نظري متطور للغاية. ومن ثم، فإن الإستراتيجية المعرفية المتبناة داخل العلم المقاد بالبيانات هي استخدام أساليب موجهة لاكتشاف المعرفة من أجل تحديد الأسئلة (الافتراضات) المحتملة الجديرة باختبارات وبحوث إضافية.

فالعملية يتم توجيهها من ناحية أن النظرية الحالية تستخدم من أجل توجيه عملية اكتشاف المعرفة، وذلك بدلاً من أن نأمل ببساطة أن نحدد جميع العلاقات داخل مجموعة البيانات وأن نفترض أنها ذات معنى بطريقة أو بأخرى. ومن ثم فإن كيفية تكوين البيانات أو إعادة تحديد الغرض منها تقوده افتراضات معينة تعززها خبرات ومعارف عملية ونظرية، مثل ما إذا كانت التقنيات وإعدادها سوف تلتقط أو تنتج مواد بحثية مفيدة وملائمة. فبدلاً من إنتاج البيانات بأي وسيلة ممكنة، وذلك باستخدام أي نوع من التقنيات المتاحة أو أي نوع من أطر التمثيل، يتم التفكير بعناية في إستراتيجيات تكوين البيانات وإعادة توظيفها، مع اتخاذ قرارات إستراتيجية من شأنها أن تؤدي إلى أنواع معينة من البيانات. وعلى نحو مشابه، يتم توجيه كيفية معالجة وإدارة وتحليل هذه البيانات من خلال الافتراضات المتعلقة بماهية الأساليب التي ربما توفر رؤى هادفة. فالبيانات لا يمكن إخضاعها لأي صياغة وجودية ممكنة، ولا يمكن إخضاعها كذلك لأي نوع من أساليب التنقيب عن البيانات، على أمل أنها ربما تكشف بعض الحقائق المخفية. وبدلاً من ذلك، يتم اتخاذ القرارات المبنية على النظرية بشأن الطريقة الأمثل للتعامل مع مجموعة البيانات حتى يمكنها أن تكشف المعلومات التي ربما تحمل فائدة محتملة وتكون جديرة ببحوث إضافية. وبدلاً من اختبار صحة أي علاقة يتم كشفها، فإن الاهتمام يتركز على تلك العلاقات -بناءً على بعض المعايير - التي يبدو أنها تؤدي إلى المسار الصحيح أو من المحتمل بشكل





أكـبر أن تؤدي إلى المســار الصحيــح. وفي الواقع، فإن الكثير من العلاقــات المفترضة داخل مجموعات البيانات يمكن تنحيتها على أنها تافهة أو ســخيفة من قبل متخصصي المجال، في حين يتم وسم البعض الآخر على أنها تستحق اهتماماً إضافياً (H.J. Miller 2010).

إن مثل عملية اتخاذ القرار هذه فيما يتعلق بطرق توليد البيانات والتحليل تعتمد على التفكير الاســتدلالي. والاستدلال هو وسيلة للاســتنتاج والتفكير المنطقي تم وضعه بواسطة C.S. Peirce (1839 – 1914) (H.J. Miller 2010). ويهدف الاستدلال إلى خاتمة تكون ذات معنى منطقي ومعقول، ولكنها ليســت حاســمة في ادعاءاتها. فعلى ســبيل المثال، لا توجد أي محاولة لاســتنتاج ما هي أفضل طريقة لتوليد البيانات، ولكن يمكن بدلاً من ذلك تحديــد طريقة ذات معنى منطقي بالنظــر إلى ما هو معروف بالفعل عن إنتاج مثل هذه البيانات. ويســتخدم الاستدلال بشكل شائع في العلوم وخاصةً في تشكيل الافتراضات، على الرغم من أن مثل هذا الاســتخدام غير معترف به على نطاق واســع. ولذا، فإن أي علاقات يتم كشــفها داخل البيانات لا تنشــأ من لا شيء، ولا تتحدث عن نفسها ببساطة كذلك. إن عملية الاستقراء – للرؤى الناشئة من البيانات – تتم صياغتها بشكل سياقي. وهذه الرؤى ليســت نهاية مطاف الاســتقصاء، إذ يتم ترتيبها وتفســيرها من خلال نظرية ما. وبدلاً من ذلك، فإن الرؤى توفر الأســاس لتشكيل الافتراضات والاختبار الاستنتاجي لصحتها. وبعبارة أخرى، يُعد العلم المقاد بالبيانات نسخة معاد تشكيلها من الطريقة العلمية التقليدية، مما يوفر وسيلة جديدة لبناء نظرية، وإضافة إلى ذلك، فإن التغيير المعرفي هو هادف بحد ذاته.

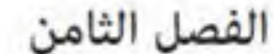
ويدعي البعض أن العلم المقاد بالبيانات سوف يصبح النموذج الجديد للطريقة العلمية في عصر البيانات الكبيرة والمتدرجة لأن التفضيل المعرفي يتناسب مع استخلاص رؤى إضافية وقيمة سيفشــل العلــم التقليدي المقاد بالمعرفة في توليدهــا (Kelling et al. 2009; H.J Miller 2012; Loikides 2010). إن العلم المقاد بالمعرفة، عند استخدامه لمنهج استنتاجي مبــاشر، له فائدة معينة في فهم وتفســير العــالم تحت ظروف البيانات النادرة والحوســبة الضعيفــة، ولكن الاســتمرار في اســتخدام مثل هــذا النهج لا يبدو أنه مفهوم بشــكل كلي عندما تشــير التطورات التقنية والمنهجية إلى أنه مــن الممكن الاضطلاع بتحليلات أكثر ثراءً للبيانات – تطبيق تحليلات جديدة للبيانات وتوفر القدرة على ربط بيانات كبيرة ومتفاوتة



سـويا ًبطرق كانت غير ممكنة لوقت قريب، والتي تقدم بيانات جديدة قيمة – والتحديد والتعامل مع الأسئلة بطرق جديدة ومؤثرة. فالعلم المقاد بالبيانات، بحسب ما يتم نقاشه، يتناسب أكثر مع استكشاف واستخلاص القيمة وفهم مجموعات البيانات الهائلة والمترابطة، وتعزيز البحوث متعـددة الاختصاصات التي تدمج خبرات المجال من خلال الإطار النظري الاسـتهلالي كونها أقل محدودية، وسـوف تؤدي إلى نماذج ونظريات أكثر شمولية وعمومية للأنظمة المعقدة بالكامل بدلاً من عناصر منها (Kelling et al. 2009).

فعلى سـبيل المثـال، يدعي البعض أن العلم المقاد بالبيانات سـوف يحـول فهمنا عن الأنظمـة البيئية (Bryant et al. 2008; Lehning et al. 2009). وسـوف يتيح إمكانية التكامل بين بيانات آنية ولحظية وعالية الدقة وذلك في سبيل تقديم نماذج تفصيلية للبيئة المتغـيرة على النحو الذي يتعارض مع نقاط ثابتة غير متغـيرة عبر المكان والزمان. ويُقصد هنـا التكامل بـين البيانات التي يتـم توليدها من العديد من المصـادر كمحطات الطقس التقليدية والمتحركة، والأقمار الصناعية والتصوير الجوي، ورادارات الطقس، ومحطات رصد التيـار، ومحطات القياس، وملاحظات المواطنين، والاستشـعار الجـوي واختبار جودة المياه ومقاييـس الغاز، وفحص التربة، والحساسـات الموزعة التي تقيـس الظواهر المختارة مثل درجـة الحرارة والرطوبة. ومن شـأن ذلك التكامل أن يُسـهم في تحديد العلاقات الخاصة بـين الظواهر والعمليـات التي تولد افتراضات ونظريات جديـدة يمكن اختبارها بعد ذلك مـن أجل التحقق من صحتها. وهذا سـوف يسـاعد أيضًا على تحديـد وفهم نقاط الربط بـين مجالات بيئية مختلفة مثل الغلاف الجوي (الجو)، والغلاف الحيوي (الأنظمة البيئية)، والغلاف المائي (أنظمة الميـاه)، والقشرة الأرضية (القشرة الصخرية للأرض)، والغلاف الترابي (التربة)، والمسـاعدة على دمج النظريات في مجاميع نظرية أكثر شـمولية. ومن شأن هذا أن يوفـر فهماً أفضل للعمليات المترابطة والمتنوعة في أرض الواقع والعلاقات الترابطية بينها وبين الأنظمة البشرية، ويمكن استخدامه كذلك من أجل توجيه النماذج والمحاكاة من أجل التنبؤ بالاتجاهات طويلة المدى والإستراتيجيات التكيفية الممكنة.





العلوم الاجتماعية الحاسوبية والعلوم الإنسانية الرقمية (Computational Social Sciences and Digital Humanities):

في حين أن معرفية العلم المقاد بالبيانات تبدو كما لو كانت وضعت لتحويل منهج البحث المطبق في العلوم المادية، والطبيعية، والحياتية، والهندسية، إلا أن مسارها في العلوم الاجتماعية والإنسانية يبدو أقل تأكيداً. فهذه المجالات من العلوم متنوعة بشكل كبير في أسسها الفلسفية، وخاصةً مع قيام عدد قليل فقط من العلماء في هذه المجالات بتوظيف المعرفة المشتركة بين العلوم. فهؤلاء الذين يستخدمون الطريقة العلمية من أجل تفسير ونمذجة الظواهر الاجتماعية، بشكل عام، يستفيدون من أفكار الفلسفة الوضعية (على الرغم من أنهم ربما لا يستخدمون مثل هذا العنوان، Kitchen 2006). ومثل هذا العمل يميل إلى التركيز على المعلومات الواقعية والكمية - الظواهر المشاهدة تجريبياً والتي يمكن قياسها بشكل فعال كالأعداد والمسافات والتكلفة والوقت على النحو الذي يتعارض مع المجالات الأكثر ماديةً من الحياة الإنسانية مثل المعتقدات والأيدولوجيات - باستخدام الاختبار الإحصائي من أجل تأسيس علاقات سببية وبناء النظريات والنماذج التنبئية والمحاكات. فالمناهج الوضعية مؤسسة بشكل جيد في الاقتصاديات، والعلوم السياسية، والجغرافيا البشرية، وعلم الاجتماع، ولكنها نادرة بشكل كبير في العلوم الإنسانية. وعلى الرغم من ذلك، فداخل هذه التخصصات المذكورة، كانت هناك حركة قوية خلال النصف الأخير من القرن الماضي تجاه المناهج ما بعد الوضعية، وخاصةً في الجغرافيا البشرية وعلم الاجتماع.

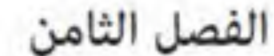
وفيما يتعلق بعلماء الوضعية في العلوم الاجتماعية، فإن البنى التحتية للبيانات، والبيانات المفتوحة والبيانات الكبيرة توفر فرصاً هائلة من أجل تطوير نماذج أكثر تطوراً وأكثر دقةً وعلى نطاق واسع للحياة الإنسانية. فثورة البيانات توفر إمكانية الانتقال من دراسات المجتمعات التي تتسم بالندرة في البيانات إلى الدراسات الثرية في البيانات، ومن اللقطات الثابتة إلى الانتشار الديناميكي، ومن التجميعات الرديئة إلى الدقة العالية، ومن النماذج البسيطة نسبياً إلى محاكاة أكثر تعقيداً وتطوراً (Kitchen 2013). إن هناك احتمالية لعصر جديد من العلوم الاجتماعية الحاسوبية التي تجري دراسات ذات اتساع، وعمق، ومدى، وإطار زمني



أكبر بكثير، وهي تغطي فترات زمنية متعددة بشكل متأصل، وذلك على النقيض من بحوث العلوم الاجتماعية الحالية (Lazer et al. 2009; Batty et al. 2012). وعلاوةً على ذلك، فإن حجم، وصحة، وتنوع، وشمولية، ودقة، وارتباطية، ومرونة، وتوسعية البيانات، بالإضافة إلى القوة المتزايدة للحوسبة وتحليلات البيانات الجديدة، تتعامل مع بعض انتقادات علماء الوضعية حتى اليوم، وخاصةً تلك المتعلقة بالاختزالية والشمولية، وذلك من خلال تقديم تحليلات أكثر دقةً وحساسيةً والتي يمكنها أن تهتم بالسياق والاحتمالية، ويمكن استخدامها من أجل صقل وتوسيع الفهم النظري للعالم الاجتماعي المكاني. وعلاوة على ذلك، وبالنظر إلى شمولية البيانات، فمن الممكن اختبار صحة مثل هذه النظرية عبر العديد من المواقف والبيئات. وفي مثل هذه الظروف، فإن المعرفة عن الأفراد، والمجتمعات، والجماعات، والبيئات سوف تصبح أكثر تبصراً ونفعاً وذلك فيما يتعلق بصياغة السياسة والتعامل مع القضايا المتعددة التي تواجه البشرية. إن التحدي الكامن في مثل هذا السيناريو لم يعد هو البيانات، ولكن تأسيس طرق إحصائية ونمذجة جديدة لها المقدرة على التعامل مع ملايين أو بلايين المشاهدات (Batty et al. 2012).

وبالنسبة لعلماء ما بعد الوضعية، فإن ثورة البيانات المنتشرة توفر كلاً من الفرص والتحديات. فالفرص تتمثل في رقمنة وربط البيانات التناظرية وغير المنظمة، والتي يُعد الكثير منها جديداً كوسائل التواصل الاجتماعي، والكثير منها كان من الصعب الوصول إليها في السابق كملايين الكتب، والمستندات، والصحف، والصور، والأعمال الفنية، والأشياء المادية... إلخ من عبر التاريخ والتي تم تحويلها إلى الصيغة الرقمية خلال العقدين السابقين من خلال مجموعة من المنظمات (Cohen 2008)؛ وإتاحة أدوات جديدة لمعالجة وإدارة وتحليل البيانات؛ وتوفر وسائل جديدة لتعزيز التعاون بين مختلف التخصصات وتجاوز نموذج العالم الواحد؛ وتوسعة الروابط البحثية مع المؤسسات غير الأكاديمية وتضمين جمهور جديد على نطاق أوسع (Sword 2008). وتخضع هذه الفرص للدراسة على نطاق واسع من خلال المجال الناشئ للإنسانيات الرقمية، والذي يجمع العلماء سوياً من جميع العلوم الإنسانية كالمؤرخين، واللغويين، والفنانين المبدعين، وعلماء الأدب والإعلام... إلخ مع أخصائي المكتبات، وعلماء الحاسب، والبيانات، والمعلومات.





وبشـكل مبدئي، فإن العلوم الإنسـانية الرقمية تتشـكل من معالجة وتحليل البيانات التي تكونت في الأصل بشـكل رقمي، ورقمنة وأرشـفة المشروعات التي تسعى إلى تحويل النصوص والمـواد التناظرية إلى صيغ رقمية بحيث يمكن تنظيمها والبحث فيها وإخضاعها لأشـكال أساسية من التحليلات الآلية، والموجهة، والشـاملة، مثل العروض المرئية الموجزة للمحتوى أو الروابط بين المستندات، والناس أو الأماكن (Schnapp and Presner 2009). ونتيجـةً لذلـك، فلقد ادعـى المؤيدون أن المجال قـد تطور لكي يوفـر أدوات أكثر تطوراً لمعالجـة، وبحث، وربط، ومشـاركة، وتحليـل البيانات التي تسـعى إلى إتمام ودعم طرق العلوم الإنسانية الحالية، وتسهيل الأشكال التقليدية من التفسير وبناء النظرية، وذلك بدلاً من اسـتبدال الطرق التقليدية أو تقديم منهج وضعي أو تجريبي لعلماء العلوم الإنسانية (Berry 2011; Manovich 2011).

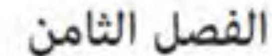
إن ما توفره العلوم الإنسـانية الرقمية هو وسـيلة لمعالجة ومشـاركة البيانات، وزيادة التحليـل التقليدي، وتعزيز البحوث عبر مدى وحجم أكبر بكثير من المصادر، وتوفير الوقت والجهد. وبدلاً من التركيز على حفنة من روايات العصر الفيكتوري أو صور من أوائل القرن العشريـن أو اثنين من فناني عـصر النهضة وأعمالهم، يصبح من الممكن البحث والربط عبر عدد كبير للغاية من الأعمال ذات الصلة، وبدلاً من التركيز على حفنة من المواقع الشـبكية أو غرف الدردشة أو الصور أو الفيديوهات أو الصحف الإلكترونية، يصبح من الممكن دراسة الآلاف مـن مثل هذه الوسـائط. ومن ثم فلقد ادعى Manovich (2011) أن: "الدراسـة المعززة بالحاسـب لمجموعات بيانات ثقافية هائلة عادةً ما تكشف أنماطاً جديدة للبيانات التي لا يمكن حتى من خلال القراءة المتأنية لأفضل دليل أن تكشفها". وبالطبع، فإن جيشاً من علماء العلوم الإنسـانية لن يكون قادراً بشـكل حريص على القراءة المتأنية لمجموعات بيانات هائلة في المقام الأول. وادعى Moretti (2005) أن العلوم الإنسانية الرقمية ستطبق الصرامة المنهجية على التخصصات التي كانت في السـابق غير نظامية وعشوائية في تركيزها ومنهجها المتبع. وادعى أن الأدبيات ذات الصلة بالأعداد، والأشـكال، والوسائط المرئية من المحتمل أن توفر رؤى جديدة وتضفي رونقاً جديداً على المجالات المنهكة.



وفي الوقـت ذاته، وكما أن هناك فرصًا، هناك أيضًا مجموعة من التحديات. فبالنسـبة للمعارضين، فإن اسـتخدام طرق العلوم الإنسـانية الرقمية الجديدة سـوف تعزز ما أطلق عليه Moretti (2005) مصطلح "القراءة عن بعد" حيث يمكن لأجهزة الحاسـب أن تعمل مثل آلات القراءة (Ramsay 2010). فمثل هذا النهج يتعارض مع الممارسـة الراسخة منذ أمـد بعيد والمتعلقة بتوفير قـراءة متأنية للمصدر. وقد ادعـى Culler (2010: 20) على سـبيل المثال بأن: "ما يتعارض مع القراءة المتأنية ليس القراءة عن بُعد ولكن شـيئاً ما مثل القراءة غير المتقنة أو القراءة العرضية". واستمر قائلاً بأن القراءة المتأنية تتضمن الاهتمام بكيفية إنتاج أو نقل المعنى، وبنوع الإسـتراتيجيات والأساليب الأدبية والبلاغية المستخدمة مـن أجـل تحقيق ما يُمكن للقارئ أن يأخـذ به على أنه المؤثر على العمـل أو الفقرة (ص 22)، وهـذا شيء لا يمكـن للقراءة عن بعد أن تقوم به، فقلـق Culler ناتج عن أن العلوم الإنسـانية الرقمية لا تسهل أو تعزز "القراءة عن بعد" فحسب، بل إنها تجعل "من الممكن إجراء البحوث الأدبية بدون القراءة على الإطلاق: مثل إيجاد كافة الحالات التي وردت فيها كلمتي "يتسول" و"متسول" في روايات كتبها اثنان من المؤلفين ثم كتابة النتائج" (ص 24).

وعلى نحو مشـابه، ذكـرت Trumpener (2009: 164) بأن: "النمـوذج المقاد إحصائياً لتاريـخ الأدب... يبـدو أنه يتطلب يداً مجهولة غير مرئية. وعـلاوةً على ذلك، فإنه يحدد الأنماط ولكنه يتجنب تحديد السـببية، والتي لا يمكن تحديدها من خلال تتبع محفوظات الناشرين، وقراءة مسـودات المخطوطات الفردية في مكتبات الكتب النادرة ومحاولة فهم، مـن كتاب لكتاب، من قام بتحديد عنوان كل رواية: المؤلف، الناشر، أو وكيل الدعاية" (ص 164). وبالنسبة لها، فإن مثل هذا البحث يتضمن حركة حقيقية وليس استخدام مجموعة من الخوارزميات. كما أكدت أن تحليل آلاف النصوص يعني "أن أي نص معين سوف يصبح غير ذي صلة من الناحية الإحصائية" (ص 164). واستمرت قائلةً " أي محاولة لرؤية الصورة بالكامل تحتاج إلى دعم من قبل المعرفة الواسعة وفهم تأريخي فطن بكيفية عمل الأساليب والمؤسسات الأدبية والأدوات التفسيرية القاطعة. وفي معرض تقديرها للجزافية "، اختتمت كلامهـا (ص ١٧٠ – ١٧١) قائلة إنه من "المهم أن معظمنا تخلى عن الاسـتمرار في البقاء في المكتبـة... فاليأس من عدم توافر المعرفة الشـاملة يجب ألّا يدفعنـا إلى التفكير المتطرف المعاكس بأنه يجب علينا أن نبدأ في معالجة الأدبيات بكميات هائلة لكي نحرز أي تقدم".





وبالمثل، فبالنسبة لـ Marche (2012)، فإن الآثار الثقافية مثل الأدبيات لا يمكن معاملتها كونها مجرد بيانات. فالجزئية الأدبية المكتوبة لا يمكن أن تكون ببساطة مجرد ترتيب للحروف والكلمات، بل هي مقترنة بالسياق وتنقل المعنى ولها سمات تفوق الوصف. فالخوارزميات ضعيفة للغاية فيما يتعلق بالتقاط وفك تشفير المعنى أو السياق، وكما ادعى Marche فإن الخوارزميات تعامل جميع الأدبيات كما لو كانت هي نفسها. واستمر قائلاً: "إن التحليل الخوارزمي للروايات والمقالات في الجرائد يكون بالضرورة ضمن حدود الاختزالية. إن عملية تحويل الأدب إلى بيانات تزيل التفوق ذاته، وهي تزيل التذوق وتزيل جميع الدقة من النقد، وتزيل التاريخ المرتبط بإطلاق هذا العمل، فرواية "إلى المنارة" ما هي إلى رواية أخرى من ضمن ركام الروايات".

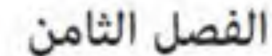
وعلاوةً على ذلك، فبالنسبة للآخرين، فبدلاً من العمل على المصادر الأصلية، فإن العلوم الإنسانية الرقمية تتعامل مع الأشياء الرقمية، التي تُنقل دفعة واحدة وتختلف من الناحية الكيفية عن أصولها (بمعنى أنها وصف للوصف) وذلك من شأنه أن يؤثر بشكل حتمي على تفسيرها.

وبالنسبة للكثيرين، فإن العلوم الإنسانية الرقمية، في هذه الحالة، تعزز التحليلات السطحية الضعيفة بدلاً من الرؤى العميقة المتخلخلة. ولذا، يتم نقدها على اختزاليتها المفرطة وعدم إتقان أساليبها، وتضحيتها بالتعقيد، والتدقيق، والسياق، والعمق، ونقدها للأنماط الوصفية، والآلية، والمدرجة، والمتسعة، والانطباع بأن التفسير لا يتطلب معرفة سياقية عميقة. ومن ثم، فلقد ادعى Jenkins (2013) أن: "قيمة الفنون وجودة المسرحية أو الرسم لا يمكن قياسها على الإطلاق. فأنت بمقدورك وضع جميع أنواع البيانات في آلة: التواريخ، والألوان، والصور، وإيصالات شباك التذاكر، ولكن لا يمكن لأيٍّ منهم أن يفسر ما هو العمل الفني، وما الذي يعنيه، وما السبب في قوته، فالإجابة عن هذه الأسئلة تتطلب إنساناً وليس آلة. وفي حين أنني متأكد أن البيانات الكبيرة مفيدة، إلا أنها لا تنطبق على جميع مجالات الحياة البشرية، فهي لن تحسن الظروف البشرية. فالحب، والثقافة، والفن، والسياسة أمور ضرورية في حياتنا ولكن البيانات الكبيرة تخبرنا القليل عن هذه الأمور. وباختصار، لا تصدق جميع الدعايات عن البيانات، فهي ليست الإجابة عن الأسئلة الكبيرة".

وعلاوةً على ذلـك، فإن التمويل المحدود المخصص للعلوم الإنسـانية تتم إعادة تركيزه على مشروعات العلوم الإنسـانية الرقمية وذلك على حسـاب الدراسات الأكثر تقليديةً مما يحرم بعض تخصصات العلم مـن التمويل. وتسري تأثيرات عمليات إعادة توزيع التمويل أيضًـا داخل العلوم الإنسـانية الرقمية ذاتها، حيث غالباً ما يتـم توجيه التمويل تجاه بناء البنى التحتية للبيانات وإنشاء أدوات جديدة، وذلك بدلاً من تمويل مشروعات تستفيد من المجموعات الرقمية من أجل الإجابة على الأسئلة الأكثر جوهريةً. والنتيجة توجيه التمويل تجاه هندسة البرمجيات، وتطوير أدوات تحليلية، ومعدات، ومختبرات، ومشروعات الموارد الأخرى، وذلك بدلاً من تمويل العلوم الإنسانية بذاتها.

ويمكن أن تنطبق ذات أنواع المناقشات هذه فيما يتعلق بالعلوم الاجتماعية الحاسوبية. فعلى سـبيل المثال، في مناقشة تتعلق بتأليف التصوير في ويكيبيديا، ادعت فيرناندا فيجاس (Frenanda B. Viegas) والتي تعمل في مختبر الاتصالات المرئية في (IBM) أن التخطيط المرئي للبيانات يمكن أن "يسـاعد الباحثين على تحديد الأسـئلة التي ربما يمكن استكشـافها من الناحية الإحصائية – على سبيل المثال، ما مدى تكرار حدوث التخريب، وما مدى سرعة ارتـداد النـص؟" (مقتبـس في Bollier 2010: 11). وعلى الرغم من ذلـك، تعد مثل هذه الأسـئلة بسيطة نسـبياً. ففي هذه الحالة، ربما يكشف التخطيط المرئي أن موضوعاً ما تتم مناقشـته بحماسة بالغة والمدى الذي يمكن أن يجعل محرري ويكبيديا يتدخلون في عملية تحريـر مدخلات البيانـات، ولكن التخطيط المرئي يفعل القليل للإجابة على الأسـئلة الأكثر نفعاً وتشـويقاً وكشـفاً، مثل أسـباب وطبيعة النزاعات والقصص البديلة التي يتم إنتاجها، ولا تتم دراسة أي منها بشكل أمثل من خلال الإحصائيات. وعلى نحو مشابه، فإن خريطة لغة التغريدات في مدينة ما ربما تكشـف أنماط التركيز الجغرافي لجماعات عرقية متعددة، ولكن الأسـئلة المهمة هي من يشكل مثل هذه التركيزات، وما السبب في وجودها، وما هي العمليات التي أسهمت في التشكيل والتكاثر، وما هي نتائجها الاجتماعية والاقتصادية. إن تحديـد الأنماط شيء؛ ولكن تفسـير هذه الأنماط شيء مختلف تمامـاً، حيث يتطلب نظرية اجتماعية ومعرفة سياقية عميقة. وفي مثل هذه الحالات، فإن الخطر يكمن في أن البحوث تمارس ببسـاطة بسبب توافر البيانات، مما يحد بشكل كبير من الأسئلة التي يمكن توجيهها وذلك لأن البيانات لم يتم إنتاجها مع وجود مثل هذه الأسئلة على البال (Vis 2013). وكما





ادعــى Gonzalez Bailom (2013): "إن تخطيــط التوزيع المكاني للمشــاعر الإيجابية، أو التكرار الذي يتم من خلاله ذكر بعض الكلمات في التواصل على الإنترنت لا يخبرنا الكثير عن توافق هذه الأنماط مع الديناميكيات الاجتماعية التي تشــكلها وتخلقها... إن قياس الأشياء بسبب أنه يمكن قياسها فحسب، لا يجعلها مشوقة أو ذات صلة".

وفي حالات مثل خريطة التغريدات، فإن الأنماط المكتشفة نادراً ما يتعين أن تكون نقطة النهايــة وهو ما يبــدو عليه الحال في بعض المشروعات (Rogers 2013). والأحرى أن تمثل الأنماط المكتشفة نقطة البداية مما يتطلب بالتأكيد أشكالاً أخرى من التحليلات وربما بعض مجموعات البيانات الأخرى.

وكــما هو الحال مع الانتقادات الأولى للعلوم الاجتماعيــة الوضعية الكمية، فإن العلوم الاجتماعيــة الحاســوبية يُــزج بها في مهام من قبل مــا بعد الوضعيين كونهــا آلية، ومفتتة ومحــددة، وتُقلص الأفراد المتنوعين، والبنــى الاجتماعية المعقدة متعددة الأبعاد، إلى مجرد نقــاط بيانات (Wyly منشــور في الصحافة). وعــلاوةً على ذلك، فإن التحليل يفســد من قبــل افتراضات الحتمية الاجتماعية كما أوضــح Pentland (2012) ذلك بالأمثلة قائلاً: "إن نوعية الشــخص الذي تكون عليه تتحدد بشــكل كبير من قبل بيئتك الاجتماعية، ولذا إذا كان بإمكاني رؤية بعض ســلوكياتك، فيمكنني أن أســتنتج الباقي، وذلك فحسب من خلال مقارنتك مع الأشــخاص في جماعتك". وعلى النقيض من ذلك، يدعي البعض أن المجتمعات البشريــة معقدة للغاية وفوضوية وغير متوقعة لكي يتــم تحويلها إلى صيغ وقوانين، حيث توفر النماذج الكمية رؤى قليلة لتفسير الظواهر مثل الحروب، والإبادة الجماعية، والعنف الأسري، والعنصريــة. وعــلى الرغم مــن أن البيانات الكمية يمكنهــا أن توضح بعض هذه المجالات، فإنها توفر رؤى شــاملة فحســب للأنظمة البشرية الأخــرى مثل الاقتصاد، وهي تفسر بشــكل غير كافٍ دور السياســة، والأيدلوجيا، والبنى الاجتماعية، والثقافة (Harvey 1972). إن النــاس لا يتصرفــون بطرق منطقية ومحددة ســلفاً، ولكــن بدلاً من ذلك فهم يعيشــون حياة مليئة بالتناقضات، والمفارقــات، والحوادث غير المتوقعة، فالكيفية التي يتم مــن خلالها تنظيم وعمل المجتمعات تختلف عبر الزمان والمكان، ولا يوجد شــكل مثالي أو أفضل أو ســمات عالمية. وفي الواقع، يوجد تنوع مدهش في الأفراد، والثقافات، وأســاليب



الحياة على الكوكب. فتقليص مثل هذا التعقيد إلى الموضوعات المجردة التي تشغل النماذج العالمية يؤدي إلى الكثير من العنف الرمزي فيما يتعلق بخلقنا للمعرفة. وعلاوةً على ذلك، فـإن المناهج الوضعية تتجاهـل عن عمد المجالات الميتافزيقية (مـا وراء الطبيعة) للحياة البشرية (المتعلقة بالمعاني، والمعتقدات، والخبرات) والأسـئلة المعيارية (المعضلات الأخلاقية والمعنوية بشـأن كيف تكون الأشـياء على النحو الذي يجب أن تكون عليه مقارنةً بالوضع الحالي لهذه الأشـياء) (Kitchen 2006). وبعبـارة أخرى، فإن المناهج الوضعية تركز فقط على أنواع معينة من الأسـئلة، والتي تسعى إلى الإجابة عنها بطريقة اختزالية تتجاهل على ما يبدو ما الذي يعنيه أن يكون المرء إنساناً وأن يعيش في مجتمعات وأماكن خصبة التنوع، وهـذا لا يعني أن المناهج الكمية غير مفيدة بل هي مفيدة بشـكل واضح للغاية، وخاصةً فيما يتعلق بتفسـير ونمذجة الأنظمة الآلية ولكن ينبغي فهم جوانب قصورها فيما يتعلق بفهم الحياة البشرية ودمجها مع المناهج الأخرى.

ومـن ثم فلقـد ادعـى Brooks (2013b) أن تحليلات البيانـات الضخمة تتصارع مع الطبيعـة الاجتماعية كـون الناس غير منطقيين ولا يتصرفون بشـكل متوقع وكون الأنظمة البشريـة معقدة بشـكل غير معقـول ولديها علاقـات متناقضة ومتباينـة. وهي تتصارع كذلك مع السـياق كون البيانات منفصلة بشكل كبير عن السياق الاجتماعي، والاقتصادي، والسـياسي، والتاريخي، وهي أيضًا تخلق أكواماً هائلة من القش كونها تتكون من ارتباطات غير طبيعية هائلة للغاية مما يجعل من الصعوبة بمكان تحديد مكان الإبرة أو ما نبحث عنه وكونها تواجه مشاكل عند مجابهة المشاكل الكبيرة وخاصةً المشاكل الاقتصادية والاجتماعية. وهي كذلك تفضل المستنسخ على الأصل كونها تحدد الاتجاهات ولكنها لا تحدد بالضرورة السـمات المهمة التي ربما تتحول إلى اتجاه وكونها تحجب القيم والمصالح المتعلقة بمنتجي البيانات وهؤلاء الذي يحللون البيانات وأهدافهم.

وهـم يفعلون ذلك، لأنه وكما قـال Peter Could (1981: 174)، وهو مؤيد للمناهج الإحصائية في تحليل البيانات: "يبدو أن رموز لغاتنا الرياضية مقيدة ومحددة بشـكل كبير. وفي الغالب هم يحدثون أضراراً ليس لأنهم يُقيدوننا فحسـب، ولكنهم في الواقع يمحون من الوجود ما يمكن أن يكون مسـتحقاً لانتباهنا واهتمامنا. وإذا ما تصورنا، وهذا ما نفعله في

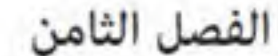


الغالب، الطبيعة متعـددة الأبعاد التي تصف تعقيد حياتنا المعاصرة... والفراغات الناتجة بواسـطة الفلتر الخطي (Strainer Called) المسـمى تحليل العامل (Factor Analysis)، فربما نقوم بسـحق المعلومات خارج الوجود وذلك تحت مسـمى التبسـيط. إنني لست متأكداً تماماً من السـبب وراء فعل ذلك، عدا أن لدي شعوراً مؤكداً تقريباً بأن توافر أجهزة حاسوب كبيرة وسريعة، وكذلك البرامج الجاهزة له يد في ذلك".

وبالنسـبة لـ Gould، فإن فهم السلوك البشري والمجتمعات البشرية لا يمكن ولا يجب تحويلـه إلى تحليلات متكررة ومنهجية وآلية، ولقد دعا إلى التفكير المسـتدام بشـأن أنواع الأسـاليب التي يتعين اسـتخدامها مع أي نوع من البيانات، وتحـت أي ظروف، وذلك من أجل الإجابة عن بعض الأسئلة المحددة، بدلاً من تشغيل البيانات من خلال مصنع "نقانق" إحصـائي والذي ينتـج لحوماً منخفضة الجودة من الألف إلى اليـاء بدلاً من تقطيع اللحوم المنتقاة.

وفيما يتجاوز الأسـلوب المنهجي والمعرفي، فإن جـزءاً من القضية هو أنه يبدو أن بعض تحليـلات البيانـات الكبـيرة تحدث دون وجود أي أسـئلة على البـال، أو أن يكون التركيز مدفوعاً من قبل تطبيق طريقة أو محتوى لمجموعة البيانات بدلاً من سـؤال محدد، أو أن تكون مجموعة البيانات مستخدمة من أجل البحث عن إجابات عن أسئلة لم يتم تصميمها لـكي يتم الإجابة عنهـا في المقام الأول. وفيما يتعلق بالأخير، لم يتم إنتاج بيانات تويتر ذات المكان المحدد جغرافياً من أجل تقديم إجابات تتعلق بالتركيز الجغرافي لمجموعات اللغة في مدينة ما والعمليات التي تقود مثل هذا الارتباط الآلي المكاني. فنحن يجب علينا ألّا نشعر بالدهشة ربما في حالة ما إذا كانت النتيجة لقطات سطحية فحسب، حتى وإن كانت لقطة رائعة، بدلاً من ذلك فإن ما يجب أن يدهشـنا هو الرؤى المتعمقة والمتغلغلة في جغرافيات الجنس، واللغة، والتكتلات، والانعزال في أماكن معينة، ولهذه الأسـباب، سـتظل دراسـات البيانات الصغيرة مفيدة.

وفي معـرض الدفـاع عن المنهج المسـتخدم من قبـل العلوم الإنسـانية الرقمية، ادعى Manovich (2011) أن الهـدف ليس اسـتبدال الخبراء البشريين بأجهزة الحاسـب، ولكن الهدف هو اسـتخدام التحليلات من أجل استكشاف مجموعات البيانات الهائلة مثل بلايين

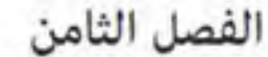




التغريدات، أو ملايين الفيديوهات على موقع يوتيوب، أو عشرات الآلاف من الروايات ومن ثـم اختيار أهداف تمثيلية معينـة من أجل التحليل اليدوي المُحكـم. وبعبارة أخرى، فإن منهج العلوم الإنسـانية الرقمية يفتح أمامنا إمكانية إجراء الدراسات السطحية على نطاق واسـع عبر الأشـخاص بدلاً من الدراسـات المتعمقة التي تركز على أعداد قليلة من الأفراد (Manovixh 2011). وبالمثـل، يدعـي البعض أن هذ المنهج يوفر العمل الدؤوب للتعمق في آلاف صفحات الأخبار ومشـاهدة مئات السـاعات من البرمجة، لـكي "يتمكن الباحثين من تركيز طاقاتهم بشـكل فعلي على التفسير" (Harris 2012)، ولذا، اقترح Flaounas et al. (2013: 111) أن: " أتمتـة الكثير من المهـام في تحليل المحتوى الإخباري لن يحل محل الحكم البشري المطلوب من أجل أشـكال التحليل النوعية والدقيقة، ولكنه سـوف يسمح للباحثـين بتركيز اهتمامهم على مـدى أبعد بكثير من أحجام عينة الأشـكال التقليدية من تحليل المحتوى. وبدلاً من قضاء الكثير من العمل الثمين على مرحلة ترميز البيانات الخام، يمكـن للمحللين التركيز على تصميـم التجارب والمقارنات من أجل اختبـار فرضياتهم، وأن يتركوا لأجهزة الحاسب مهمة اكتشاف جميع المقالات في موضوع معين، وقياس العديد من سـمات المحتوى مثل قابليتهم للقراءة، واستخدام أشـكال معينة من اللغة والمصادر، إلخ. مجرد عدد قليل من المهام التي يمكن أتمتتها في الوقت الراهن".

وبالمثـل، فلقـد ادعى Selisker (2012) أن العلوم الإنسـانية الرقمية لا تسـعى إلى "أن تسـتبدل الأدب برموز الحاسـب، أو أن تقايض التعقيد الأدبي بصيغ لا هدف لها. وبدلاً من ذلك، فإن هذه المشروعات تفكر فحسب بطرق إبداعية عن المشاكل التاريخية الأدبية التي لا يمكن التعامل معها بسـهولة بدون الحواسيب". وبالنسبة لـ Ramsay (2003; 2010)، يمكن أن تساعد الحوسبة في عملية التفسير، ولقد ادعى أن القراءة المتأنية تتضمن ممارسات اختيار وعـزل وملاحظة الأنمـاط في النصوص، وأن الترميز ذو كفاءة عالية في مثل هذه الممارسـات، وعـلاوةً على ذلك، فإن الترميز يقوم بالأمر نفسـه بطرق أكثر موضوعيةً واتسـاقاً من النظرة الذاتية للعالم، وبدلاً من استبدال التعرف على المعنى، فبالنسبة له فإن توظيف الخوارزميات يسهل هذه العملية.





إن دفاع العلوم الاجتماعية الحاسوبية أمر مختلف تماماً في الغالب، ففي حين أن جميع العلوم الإنسانية الرقمية تقر بقيمة القراءة المتأنية، وتؤكد أن القراءة عن بعد تكملها من خلال توفير التعمق والسياقية، فإن الأساليب الموضوعية للعلوم الاجتماعية تتعارض مع المناهج بعد الوضعية. إن الفرق بين العلوم الاجتماعية الحاسوبية والعلوم الإنسانية الرقمية بهذا الصدد يرجع إلى أن الإحصائيات المستخدمة في العلوم الإنسانية الرقمية وصفية بشكل كبير تحدد الأنماط وتصوغهم كأعداد وصور وخرائط. وعلى النقيض من ذلك، فإن العلوم الاجتماعية الحاسوبية تستخدم المنهج العلمي حيث تكمل الإحصائيات الوصفية بالإحصائيات الاستدلالية التي تسعى إلى تحديد السببية. وبعبارة أخرى، فإنها تحصل على الدعم من قبل المعرفة حيث يكون الهدف هو تقديم نماذج إحصائية متطورة تفسر وتحاكي وتتنبأ بالحياة البشرية. ومن الصعوبة بمكان التوفيق بين هذا وبين مناهج ما بعد الوضعية، فالدفاع في هذه الحالة يرتكز على فائدة وقيمة الطريقة والمناهج، وليس على توفير تحليل تكميلي لمجموعة من البيانات أكثر تمدداً.

وهناك بدائل لمثل هذا الموقف، مثل البديل الذي تم تبنيه في نظم المعلومات الجغرافية (GIS) والإحصائيات الجذرية (Radical Statistics)، وهؤلاء الذين يستخدمون مناهج ذات طرق مختلطة، وهذا يتضمن إما توظيف النماذج والإحصائيات الاستدلالية شريطة أن يكون هنالك إدراك لمواطن الضعف فيهما، أو، كما يحدث على نحو شائع، استخدام الإحصائيات الوصفية التي تدعمها دراسات البيانات الصغيرة. وهنا، يتم استخدام كلٍّ من البيانات الكبيرة والبيانات الصغيرة من أجل توفير رؤى مختلفة ولكن تكميلية. وبغض النظر عن مثل هذه التكميلية، فإن الكثير من علماء ما بعد الوضعية سوف يستمرون في مقاومة البيانات الكبيرة والمناهج الإحصائية بأشكالها المتعددة، وبلا شك فإن المجادلات سوف تتسارع وتيرتها في المستقبل المنظور.



الخاتمة:

هناك شـكوك قليلة بأن ثورة البيانات أتاحت احتمالية إعادة التشـكيل المعرفي للعلوم، والعلوم الاجتماعية والإنسـانية، وأن مثل إعادة التشـكيل هذه تحدث بالفعل بشكل فعال ونشط. فالبيانات الكبيرة وتوسع البيانات الصغيرة من خلال البنى التحتية للبيانات يساعدان على تنفيذ مناهج جديدة تجاه إنتاج وتحليل البيانات، مما يجعل من الممكن توجيه الأسئلة والإجابـة عنها بطرق جديدة. وقد شـجع هذا البعض على الادعاء بشـكل جريء بأن ثورة البيانـات هي ابتكار تدميري يعلن عن نقلة نوعيـة في كيفية إجراء البحوث، وهذه بالقطع هـي الحالة حيث إن التحول من صحاري البيانات إلى طوفـان البيانات وتحليلات البيانات الجديدة سـيغير الممارسـات البحثية. وعلى الرغم من ذلك، يتعين أخذ الحيطة عند القيام بتأكيدات تتعلق بإعادة تشكيل البحث وبأفضل طريق يمكن اتباعه في إعادة التشكيل هذه.

إن العلم يُعامل بشـكل سـيئ من قبل الادعاءات القائلة بأن ثـورة البيانات تعني "أن البيانات يمكنها أن تتحدث عن نفسـها" أو أنه يمكنها بدون أي مشـاكل استخدام أساليب مثل تجريف البيانات حيث يسـبق الارتباط السببية، (انظر الفصل التاسع). وتُعد المناهج التـي يتم تطويرها داخل العلوم المقادة بالبيانات والتي تمزج بين الاسـتدلال، والاسـتنتاج والاسـتقراء أكثر إنتاجية بكثير، حيث تتم إعادة تشـكيل الطريقة العلمية، وذلك بدلاً من تدميرها، إن الأسس الفلسفية للعلم المقاد بالبيانات، على الرغم من ذلك، لا تزال في أطوارها التمهيدية وهناك حاجة للتفكير المستفيض في معتقداتها المعرفية ومبادئها ومنهجيتها.

أما الموقف فيما يتعلق بالعلوم الاجتماعية والعلوم الإنسانية فهو أكثر تعقيداً بشكل ما وذلك إذا وضعنا في الاعتبار تنوع الأسس الفلسفية لهذه العلوم، حيث إنه من غير المحتمل أن تقود ثورة البيانات إلى تأسيس نموذج جديد. وبدلاً من ذلك، فإن البنى التحتية للبيانات وثورة البيانات سوف يعززان مجموعة البيانات المتوافرة للتحليل ويمكنان مناهج وأساليب جديـدة، ولكنهـا لن تحل محل دراسـات البيانات الصغيرة. وفيما يتعلـق بالعلوم، هناك حاجة لتفكير نقدي أوسـع في الآثار المعرفية لثورة البيانات على العلوم الاجتماعية والعلوم الإنسانية، وهذه المهمة بدأ بالكاد التعامل معها وذلك على الرغم من سرعة التغيير الحاصل في مشهد البيانات.


المكتبة الاقتصادية
ECONLIBRARY
قناة المكتبة على التيليجرام


المكتبة الاقتصادية



الفصل التاسع

القضايا الفنية والتنظيمية

(Technical and Organizational Issues)

كما ناقشنا في الفصلين السابقين، توفِّر البنى التحتية للبيانات ومبادرات البيانات المفتوحة وإنتاج البيانات الكبيرة فرصاً جديدة لقياس وفهم العالم. وهي أيضًا تثير قضايا فنية وتنظيمية، وتطرح أسئلة معيارية وأخلاقية. وفي هذا الفصل، سيتم التركيز على الأمر الأول، بما يعني التركيز على بعض القضايا مثل مجال مجموعات البيانات، والوصول إلى البيانات، وجودة البيانات، وتكامل البيانات، والتبادلية (القابلية للتبادل)، وتطبيق التحليلات والمغالطات البيئية، والمهارات والقدرات والإمكانات التنظيمية. فبعض هذه القضايا يمكن التعامل معها من خلال الحلول الإدارية والفنية، في حين أن البعض الآخر من هذه القضايا تمثل مشاكل عسيرة الحل ومن الصعب التعامل معها. وعلى الرغم من ذلك، فإذا كانت القضايا المتعلقة بجودة البيانات والمغالطات البيئية لا يمكن التعامل معها بشكل كاف، ففي هذه الحالة لا يهم ماهية النظريات المعرفية التي يتم تطبيقها، (انظر الفصل الثامن)، لأن التحليلات والتفسيرات ستكون محل ارتياب. وعلاوة على ذلك، فإذا كان الوصول إلى البيانات مقيداً بشكل كبير، أو إذا كانت الحكومات والشركات والمجتمع المدني يفتقرون المهارات الكافية لإدراك المغزى من هذه البيانات، ففي هذه الحالة ستكون قيمة ومنافع البيانات المفتوحة والكبيرة المتدرجة مقيدة بشكل كبير للغاية. ونتيجة لذلك، فإن القضايا التي تمت مناقشتها في هذا الفصل لا يمكن تجاهلها أو إبعادها بسهولة، ومن ثم فهي تتطلب الاهتمام البالغ. وفي الفصل التالي، سيتحول التركيز إلى قضايا سياسية واجتماعية أعم وأشمل والتي تتعلق بإنشاء ومشاركة وتحليل واستخدام البيانات مع الوضع في الاعتبار الأبعاد الأخلاقية، والقانونية، والأمنية لثورة البيانات.



الصحاري والفيضانات (Deserts and Deluges):

في كافة النقاشات الدائرة بشأن ثورة البيانات يكون في الغالب من الصعب تحديد مصدر مجموعة مناسبة من البيانات فيما يتعلق بالكثير من الظواهر، وبخاصة تلك البيانات التي تغطي فترات زمنية طويلة ومتأصلة مكانياً. وبشكل جزئي، فإن هذه تعد قضية من قضايا التغطية من ناحية، ومن قضايا الوصول من ناحية أخرى، فمن جانب التغطية، فإن جميع مصادر البيانات محددة من ناحية العينة، والمدى الجغرافي، والوقت، ومتغيرات الصفات (الحقول) والمؤشرات. وفي الواقع، فإن العالم معقد بشكل كبير لغاية أنه يستحيل التقاط نطاق كامل وكافة فروقاته الدقيقة، وتناقضاته، وتبايناته.

إن البيانات التي يتم جمعها من خلال دراسات البيانات الصغيرة تعد عينات بشكل صريح للغاية، وهي تُجمع بشكل غير مستمر، وهي ذات رقم محدد من المتغيرات وذلك بالنظر إلى تركيزها المحكم والتكلفة المرتبطة بالجمع والتحليل، فمعظم هذه الدراسات تمثل دراسات أو تجارب منفصلة تركز على قضايا وحالات معينة لمجموعة، مؤسسة، أنواع، أماكن، تقنيات... إلخ، وهي تفتقر إلى المكون الطولي حيث إنها لا تُغطي فترات زمنية متعددة وطويلة. فالدراسات التي تقوم بدراسة التغييرات بمرور الوقت عادةً ما يتم تكرارها على فترات محددة. وفي هذه الحالة، فإن فهمنا لظاهرة معينة عادةً ما يعتمد على مجموعة مقيدة من البيانات، وحتى عندما يتم تجميع وترقية مجموعات البيانات داخل البنى التحتية للبيانات، فإن التغطية تكون جزئية وانتقائية. فعلى سبيل المثال، فإن مجموعات بيانات الإدارة العامة الأيرلندية ذات تغطية محدودة وهي مملوءة بالفجوات والسكنات عن قطاعات معينة مثل قطاعات الإسكان والنقل، وغالباً ما يتم الاحتفاظ بها داخل المؤسسات (Kichin et al. 2007).

وعلى نحو مشابه، فإن البيانات الكبيرة ربما تميل إلى أن تكون بيانات شاملة، ولكن وكما هو الحال مع مجموعات البيانات الأخرى، فإن كلتيهما تكون تمثيلاً وعينة، وهي تحل، بشكل جزئي فحسب، قضية صحراء البيانات التي يقصد بها أن تكون البيانات موزعة على مجاميع بيانات متفرقة يصعب استقاء البيانات منها أو إضافة بيانات اليها. وكما لاحظنا في الفصل الثاني، فإن البيانات التي يتم التقاطها تتشكل من قبل التقنية المستخدمة، والسياق



الذي تم فيه إنشــاء هذه البيانات، وتوصيف البيانات الذي تم توظيفه. فالبيانات المنشئة تمثل عادةً ما يمكن التقاطه إلى حد ما بطريقة مباشرة، وهي تكون في غالب الأحيان منتجاً فرعيــاً للنظام، ويكون مجالها الفوري مقيداً من قبل النطاق كسلاســل الأســواق المركزية، وشــبكات المستشــعرات، ومواقع التواصل الاجتماعي. وفيما يتعلق بالعلوم، فإن البيانات الكبيرة يتم إنشــاؤها حالياً لما يتعلق بمجموعة منتقــاة من الموضوعات فقط، وهي بعيدة كل البعــد عن أن يتم إنتاجها عبر جميــع النطاقات والبؤر العلمية. وعلاوة على ذلك، فإن القرار فيما يتعلق بإنشاء البيانات الكبيرة لم يتشكل بعد، ولكنه يتطور ويتركز من أجل توفير مجموعة من الرؤى، بحيث تشــكل كل واحدة منها بعضاً مــن العناصر التي يتم التقاطها كالاتســاع، والعمق، والتفاصيل... إلخ. وحتى عندما يتم دمج بيانات كبيرة مع مجموعات البيانــات الأخــرى، فإن مجالهم يكون بعيــداً كل البعد عن كونه مجالاً عالميــاً. ولذا، فإن البيانات الكبيرة، في هذه الحالة، لا تعالج مسألة صحاري البيانات وذلك فيما يتعلق بالكثير من مجالات الدراسة.

وفي الوقت نفســه، وفي حين أن صحــاري البيانات مازال تواجدها مســتمراً، فإن البنى التحتية للبيانات والبيانات الكبيرة أنشأا طوفاناً من البيانات في بعض المجالات. وكما لاحظنا في الفصــل الرابع، كان هناك نموٌّ ملحوظٌ في حجم وسرعة إنتاج البيانات والتي تم تحديدها من أجل أن تنمو بسرعة عالية، وذلك مع إنشاء وبدء المصادر الجديدة، وتعقد مجموعات البيانات التي تم إنتاجها، والمقدرة على ربط البيانات بعضها البعض. ومن شأن هذا الطوفان أن يفرض عدداً من التحديات فيما يتعلق بالمناولة، والمعالجة، والتأمين، والتصنيف، والفهم لمثل هذه البيانات (Gantz and Reinsel 2011). وبدلاً من مجموعات البيانات الصغيرة وغير المستمرة، يحاول المحللون الاستفادة من الكميات الهائلة من البيانات التي يتم تدفقها من أنبوب ســميك. فمعظم هذه البيانات أصبحت بيانات عابرة وتم فقدها، ولم يتم أبداً تحليلها أو الاســتفادة منها. وبعض هذه البيانات قد تم تخزينها في مســتودعات البيانات التي تمثل إلى حد كبير مكبات غير مراقبة للنفايات (Asay 2013)، وهناك يتم الإبقاء عليها حتى تأتي أوقات يمكن لهذه البيانات أن تقدم فيها معلومات مفيدة. ونتيجةً لذلك، وبينما تزداد نســبة البيانات بشكل تصاعدي، فإن نسبة البيانات التي يتم معالجتها وتحليلها تقل بشــكل كبير (Zikopoulos et al 2012). وقد أشــار (2012) Wu إلى أنه في الوقت التي



تــزداد فيه البيانات، يتم الحصول على معلومات أقل كنســبة من هــذه البيانات. وبعبارة أخرى، تزداد ضوضاء البيانات بشكل كبير للغاية ولكن إشارات المعلومات تنمو بشكل أقل بكثير، وعلاوةً على ذلك، يصبح الأمر أكثر صعوبة لعزل هذه البيانات في خضم هذه الضواء (انظــر Silver 2012). وهــذا بلا جــدال وظيفة البيانات الكبيرة وعلــم البيانات حيث لا يزالان في أطوارهما الأولية، ومع الاستغلال الكامل للأساليب الحديثة والتي بإمكانها مراقبة ومعالجــة وتحليل كميات هائلة من البيانات بشــكل آني أو لحظي إلا أن هذا من المتعذر تحقيقه بشكل كبير. وحتى مع نضوج وتطور تحليلات البيانات الكبيرة، فإنه من المرجح أن تعاني من أجل تطوير طرق ومناهج يمكن من خلالها استخلاص التحليلات المفصلة وليست التحليلات الســطحية؛ والتعامل مع أسئلة "لماذا" وكذلك أســئلة "ماذا"؛ وتقديم إيضاحات وليس شروحات (انظــر الفصل الثامن). وعلاوة على ذلك، يتعين تنويع هذه الطرق وذلك من أجل مجاراة الأشكال الجديدة من البيانات التي يتم إنشاؤها من قبل المصادر المبتكرة. وحتى مع تحقيق ذلك، فإن مجاراة طوفان البيانات واســتخلاص إشــارات ذات مغزى من الضوضاء سيكون كفاحاً مستمراً.

الوصول (Access):

إن العائق الأســاسي الآخر أمام الاســتعانة بمصادر مجموعات مناسبة للبيانات هو مدى قابليــة الوصــول إلى هذه المجموعات وقيــود التراخيص التي تحدد معايير اســتخدام تلك المجموعات. وبشــكل واضح، وكما ناقشــنا في الفصلين الثاني والثالــث، فإن الوصول يعتبر قضية أساســية فيما يتعلق بالبحــوث والبيانات الممولة من قبل المؤسســات العامة والتي تنشــئها الهيئات العامة، فمحفوظات البيانــات والبنى التحتية للبيانات ومبادرات البيانات المفتوحة تُعني وتهتم بشكل كبير بمشاركة وإنتاج البيانات المتاحة للتحليل. وكما أوضحت المناقشــات في الفصول المشار إليها سابقاً، فعلى الرغم من حقيقة أن البيانات غير تنافسية، وغير قابلة للاستثناء أو الإقصاء، ولها تكلفة هامشية عند استنساخها (Floridi 2010)، إلا أن البيانات مقيدة بشكل شائع وذلك عند توزيعها. وفي بعض الأحيان، فإن الوصول المقيد إلى البيانات يكون أمراً مرغوباً فيه، وذلك إذا وضعنا في اعتبارنا طبيعتها الحساسية والشخصية وكذلك البعد الأخلاقي المرتبط بنشر واســتخدام هذه البيانات (انظر الفصل العاشر). وفي



أحيــان أخرى، يكون الوصول إلى البيانات مقيداً لأجل التأكد من أن هؤلاء الذين يحتفظون بهذه البيانات ســيرفعون من قيمة هذه البيانات ويضيفون ميزات تنافسية لها أو يحققون أموالاً من خلال البيع او الترخيص للوصول لهذه البيانات (انظر الفصل السابع). وفي أحيان أخــرى، ربمــا ترغب إحدى الهيئات في تقييد توزيع البيانــات نتيجة أنها تخشى ما يُمكن أن تكشــفه هذه البيانات، والذي ربما يكون له تداعيات سياســية أو اقتصادية، أو يكون مرد ذلك الخوف من كشف تدني جودة البيانات التي كانت محفوظة.

وفي حين أن هذه المعوقات بدأت تتلاشى ببطء فيما يتعلق بالبيانات الممولة والمستحدثة من قبل المؤسســات العامة في الدول الديمقراطيــة الغربية، إلا أن مثل هذه التحركات تعد أكثر تقلباً مع تلك البيانات المســتحدثة من قبل المؤسســات الخاصة. وفي الواقع، فإن هذا الأمــر يعد متناقضاً إلى حد مــا، إذ إنه على الرغم من طوفان البيانات الكبيرة الناشــئ، إلا أن الوصــول إلى مثل هــذه البيانات يظل مقيداً ومحدوداً في الوقــت الحالي، وذلك بوجود عــدد محــدود للغاية من الكيانــات الغارقة في مثل هــذا الطوفان والتــي يمكنها الوصول إليــه (Crawford 2012, King 2011). فالشركات من أمثال مشــغلي الهواتف الخلوية، ومطوري التطبيقات، ومزودي وســائل التواصل الاجتماعي، والمؤسســات المالية، وسلاســل التجزئــة، وشركات المراقبــة والشركات الأمنيــة، لا يقع على كواهلها أي التزامات لمشــاركة البيانات التي يســتحدثونها من خلال عملياتهم التشــغيلية بحرية. وعادةً ما يتم التفاوض للوصول إلى البيانات بشكل فردي وهذا يتضمن توقيع العديد من اتفاقيات حماية الملكية الفكريــة واتفاقيات عدم الإفصاح، وكذلك الموافقة عــلى الشروط والبنود التي تحدد كيف وفي أي غرض ســيتم اســتخدام أو لا يتم استخدام مثل هذه البيانات. وحتى بعد كل هذا، فمــن المحتمل أن تمثل البيانات التي يتم توفيرها عينة محدودة، كما من الممكن أن تفتقر هذه البيانات إلى المعلومات السياقية مثل الكيفية التي تم تعيين هذه البيانات من خلالها، أو تكون البيانات غير معالجة بشــكل كبير أي لم يتم تنقيتها أو هيكلتها، أو تكون مُجمعة، أو مُجــردة مــن المتغيرات المفيــدة، أو معالجة بشــكل أو بآخر، وذلك من شــأنه أن يُقيد أســاليب تحليل هذه البيانات. وفي بعض الحالات، يمكن أن يتم إتاحة كمية محدودة من البيانــات للباحثين والجمهور وذلك من خلال واجهات برمجــة التطبيقات (Application Programming Interfaces -APIs). فواجهــات برمجة التطبيقات تتكون من مجموعة



من الأوامر التي يمكن اســتخدامها من أجل استعادة البيانات المخزنة في قواعد بيانات على خادم متاح الوصول إليه للجمهور (Manovich 2011). فعلى ســبيل المثال، يســمح تويتر لعــدد محدود من الــشركات بالوصول إلى كافة بياناته المتدفقة عــبر موقعه، وذلك مقابل رســوم أو لأغــراض تجارية. وهو يتمتع بالصلاحيات التي تســمح لــه بإملاء شروطه فيما يتعلق بالأغراض التي يمكن فيها اســتخدام مثل هذه البيانات. فالباحثون يتم تقســيمهم إلى قســمين: القســم الأول يمكنه الوصول إلى ١٠٪ من التغريدات العامة، والقسم الثاني لا يمكنه الوصول إلا إلى ١٪ فحسب من التغريدات العامة، أو يمكنهم الوصول إلى مجموعات فرعيــة مختلفة من المحتوى (الحســابات المدرجة في القائمة البيضاء)، وذلك مع اســتبعاد التغريــدات الخاصــة والمحمية في جميــع الأحــوال (boyd and Crawford 2012). وفي حال المواقع الإلكترونية التي تحوي على ســبيل المثال قوائم منتجات وتقييماتها (Product Listings and Reviews) وأدلة أعمال (Business Directories)، وكذلك في حال وســائل التواصل الاجتماعي والتي تتضمن الملفات الشــخصية العامة (Public Profiles) والتغذية (Feeds)، فإن بعض الشركات مثــل (www.80legs.com) (80legs) وشركة ((Mozenda www.mozenda.com)) توفر خدمات زاحف الشبكة (أي تصفح الشبكة العالمية بطريقة منهجية وآليه ومنظمة) وتجريف الشــبكة (أي اســتخراج البيانات من مواقع الإنترانت عن طريــق برامج مخصصة) لحصد بيانات مفصلة من هذه المواقع أو للوصول لبيانات تغذية مبنية مسبقاً.

وما يقلق في هذا الأمر هو أن المعلومات التي يمكن أن توفرها البيانات الكبيرة المملوكة للشركات الخاصة والتي تم بيعها بشكل تجاري سوف تكون محصورة على قطاع الأعمال فحسب، أو ربما يتم إتاحتها لمجموعة مميزة من الباحثين الأكاديميين الذين لا يمكن تكرار النتائج التي توصلوا إليها أو التحقق منها (Lazer et al. 2009; King 2011). ولا أحد من هــذه المواقف يخدم المصلحة العامة، ولكن إذا وضعنــا في اعتبارنا القيمة التجارية للبيانات محل المناقشــة فإن زيادة الوصول إلى هذه البيانات لن يكون مهمة سهلة أبداً. فحقيقة أن البيانات الكبــيرة الثرية من الناحيتين الاجتماعية والثقافية مملوكة للشركات الخاصة تُشير إلى أنه في الوقت الحالي لن يكون من السهل في الواقع إجراء بحوث العلوم الاجتماعية الحسابية والعلوم الإنسانية الرقمية (Manovich 2011). فكما ذكر كلاً من



فــإن boyd and Crawford (2012: 673) " الكثــير من الحماس الذي يحيط بالبيانات الكبيرة ينبع من الإدراك القائل بأن هذه البيانات توفر وصولاً ســهلاً لكميات هائلة من البيانات. ولكن من يتمتع بمثل هذه القدرة للوصول إلى هذه البيانات؟ ولأي غرض يتم الوصول؟ وفي أي ســياق يجرى هذا الأمر؟ وما معوقات الوصول؟" فهذه أســئلة تتطلب الكثير من التفكير مع اتخاذ إجراءات مناسبة.

جودة وصحة وأصل البيانات (Data Quality, Veracity and Lineage):

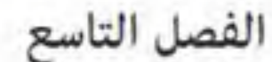
جودة وصحة البيانات هما أمران مهمان للغاية كونهما يعززان المدى الذي يمكن أن يثق فيه المرء بالنتائج المســتخلصة من تحليل البيانــات. وترتبط جودة البيانات بمدى تنظيمها أي خلوها من الأخطاء والفراغات، ونقائها أي خلوها من التحيز، وتماسكها بمعنى أن تحوي القليل من التناقضات. بينما يُشــير مصطلح صحة البيانات إلى موثوقية البيانات وإلى مدى تمثيلها للمعنى الذي تُشــير اليه بدقة (إتقان) وأمانة (إخلاص). ولهذا فإن تأسيس وتوثيق جودة وصحة البيانات يعتبر شقاً رئيسياً في سُلاَلَة أصل البيانات. فسُلاَلَة أصل البيانات هي:

المعلومــات التي تصف مصدر الملاحظات، والمنهجيات التــي خضعت لها البيانات عند جمعها وتصنيفها، والتعديلات، والتحولات، والتحليل، والاشــتقاقات التي تعرضت لها، وهي توفــر أيضًــا الافتراضات والمعايير التــي تم تطبيقها عند أي مرحلة مــن مراحل حركة هذه البيانــات، وكذلــك أي تحيزات... وتوفر سُــلاَلَة أصل البيانات مجموعــة بيانات مع أصلها وتسمح للمستخدم أن يحدد مدى ملاءمة البيانات للاستخدام (Lauriault 2012).

وهي أيضًا تصف بالتفصيل منشــأ البيانات، ومن شــأن الجودة والصحة وسُلاَلَة الأصل الضعيفــة للبيانات أن يقوضْــنَ درجة الثقة والمصداقية التي يمكن أن يضعها الشــخص في مجموعة البيانات والتحليلات التي اعتمدت على هذه البيانات.

ومــن الناحيــة التقليدية، فإن الدراســات القائمة على بيانات نــادرة والتي يكون حجم العينــات فيها صغــيراً في الغالب، فإن جــودة وصحة وأصل البيانات تكــون على قدر كبير من الأهمية. فإذا كان الشــخص سيعمل على اســتخلاص نتائج عن السكان بالكامل وذلك اعتــماداً على عينة مثالية صغيرة، فإنه يتعين في هذه الحالة أن تكون العينة واضحة للغاية





وخالية من العيوب بأكبر قدر ممكن، ويتعين أن تكون على درجة عالية من الدقة والصحة. وعلى الرغم من ذلك، فإن عدد مجموعات البيانات يتم تشكيلها بشكل تام محدود للغاية، ويمكن استخدامها بدون أي شكل من أشكال تنظيف البيانات، بمعنى التحقق من الأخطاء وتصحيحها، وإضافة البيانات المفقودة، ومعايرة وإعادة تهيئة البيانات حتى يمكن ربطها مع بيانات أخرى أو لتكون هذه البيانات سهلة الاستجابة لمناهج تحليل معينة (انظر الفصل السادس). ولذا، فإنه يتم استهلاك الكثير من العمل خلال تصميم البحث من أجل تقييد التحيزات المنهجية والتحيزات المرتبطة بتحديد العينات كي يتم التأكد من أن البيانات دقيقة وذات مصداقية بقدر الإمكان قبل تحليل أو مشاركة هذه البيانات. وتعمل المحفوظات الأرشيفية والبنى التحتية بشكل نشط من أجل التأكد من أن تظل مثل هذه البيانات ذات مصداقية ودقيقة وصحيحة بمرور الوقت، مع المحافظة على سُلاَلَة أصل البيانات من خلال الاستقرار المؤسسي، وعمليات إدارة البيانات وواصفات البيانات، وتمكين نقل البيانات عبر مختلف الصيغات والمنصات بما يحقق المؤامة بين الأجيال المختلفة من التقنيات (Lauriault 2012). وبدون مثل هذا العمل، فإن صحة البيانات المتضمنة داخل تلك البيانات سوف تتلاشى في القريب العاجل وذلك مع انعدام الثقة.

وعلى النقيض من ذلك، ادعى البعض أن دراسات البيانات الكبيرة ليست بحاجة إلى تطبيق معايير جودة وصحة البيانات ذاتها، وذلك يرجع إلى أن الطبيعة الكلية لمجموعات البيانات الكبيرة "ن= الجميع" (n=all) تجعلها تتخلص من التحيزات المتعلقة بتحديد العينات، وتقوم بالتعويض، أكثر من المطلوب، عن أي أخطاء أو ثغرات أو تناقضات في البيانات أو أي ضعف في صحة البيانات (Mayer-Schonberger and Cukier 2013). فالحجة التي استندت إليها وجهة النظر هذه أنه "مع وجود عدد قليل من الأخطاء المرتبطة بتحديد العينة، يمكننا تقبل أخطاء أكبر مرتبطة بالقياس" (ص 13)، "ويمكننا التسامح مع عدم الدقة" (ص 16). إن الثمن المقابل لـ "ن= الجميع" (n=all) هو قبول الفوضى التي نشأت بشكل كبير من جراء صعوبة تنظيف البيانات ذات السرعة العالية وغير المنظمة، وتقبل البيانات كونها "محتملة أكثر من كونها دقيقة" (ص 35). وعند النظر إلى الأمر بهذه الطريقة، ادعى (Mayer-Schonberger and Cukier 2013) أنه " كلما كان هنالك بيانات أكثر، كان ذلك أفضل" أو للتعبير عن هذا الأمر بشكل آخر، "إذا كان لديك بيانات



كثيرة جداً فإن ما هو جيد بدرجة كافية يُعد جيداً بدرجة كافية" (Helland 2011). ومن ثــم فقد ادعى Franks (2012: 211) أن المطلوب هو "بيانات نظيفة بدرجة كافية"، وهي البيانــات التي تمكن المحللين من "الحصول على النتائج التي يمكنهم الوثوق بها". وبالطبع فإن هذا يفترض أن جميع اســتخدامات البيانات الكبيرة ســوف تتسامح مع عدم الدقة، في حين أن الواقع يشــير إلى أن الكثير من اســتخدامات البيانات الكبيرة تتطلب الدقة بشــكل كبير، أو على الأقل بيانات ذات معايير الأخطاء التي يمكن حسابها أي بيانات موثوقة. فعلى ســبيل المثال، فإن موقع خريطة الطريق المفتوح (OpenStreetMap) يهدف إلى أن يكون في مستوى منتجات الخرائط الاحترافية من حيث التفصيل والدقة، ويهدف كذلك إلى كسب ثقة مســتخدميه كونه مصدراً من مصادر المعلومات. وعلى النحو ذاته، تهدف موســوعة ويكيبيديا (Wikipedia) إلى أن تكون موســوعة دقيقة وموثوقة مثل الموسوعات الأخرى. وبطريقة مماثلة يتعين أن تكون البيانات المالية والبيانات الطبية دقيقة وصحيحة من أجل أن تغرس الثقة في نفوس التجار والأطباء والمرضى... والقائمة تطول. فمثل وجهة النظر هذه تفترض أيضًا أن سُلاَلَة أصل البيانات وقضايا صحة وموثوقية البيانات ليست مهمة بالنسبة لمستخدمي البيانات.

وبغض النظر عن تســامح بعض مجموعات بيانــات البيانات الكبيرة تجاه عدم الدقة، إلا أن تحذير المعطيات الرديئة تؤدي إلى مخرجات رديئة لا يزال ســارياً. فمجموعات البيانات الكبيرة التي تنشــئ بيانات معيبة ورديئة، وتم التلاعب بها ومتحيزة، أو تنشــئ بيانات ذات مصداقية ضعيفة، ســوف تــؤدي إلى تحليلات ونتائج ذات مصداقية ضعيفة وســوف تكون الفوائد المرجوة منها قليلة للغاية بالنســبة لهؤلاء الذين يحللون ويســتفيدون منها. ويمكن للبيانــات الكبــيرة، عن طريق طــرق إنتاجها، أن تعاني من جميع العلــل، فالبيانات يمكن أن تكــون معيبة ورديئة من خلال أخطاء الأداة، حتــى عبر التقنيات التي يُفترض أن تكون ذات دقــة عالية للغاية مثل نظام التموضع العالمي (GPS) والــذي تختلف دقته من جهاز لآخر، وعدد الأقمار الصناعية المتوافرة في وقت معين من الأوقات (Dodge and Kitchin 2013). أو من الممكن أن يكون قد تم التلاعب أو التزوير في البيانات من خلال حســابات مزورة أو القرصنة من أجل التأثير على الاتجاه العام وتوجيه العامة نحو مسارات محددة (Vis 2013). فعلى ســبيل المثال، هناك مئات الآلاف من الحســابات المزيفة على تويتر، وهي تشكل ٤.c٪



من إجمالي حسابات المستخدمين على تويتر (De Micheli and Stroppa 2013)، وأيضًا فإن حتى ٤٥٪ من حسابات بعض العلامات التجارية لمتابعي تويتر هي حسابات آلية (Calzolari 2012)، وأكثر من درزنين من الخدمات تبيع حسابات تويتر مزيفة (Baym 2013).

فالبيانات يُمكن أن تتسم بالتحيز جراء بعض الأسباب. إن الخصائص السكانية التي يتم اختيارها في العينات نادراً ما تكون شاملة وتتشوه إحصائياً من قبل الجنس، والنوع، والدخل، والموقع، والعديد من العوامل الاجتماعية والاقتصادية الأخرى. فليس جميع الأشخاص يستخدمون تويتر أو الفيسبوك أو يتسوقون من محلات معينة، أو يكونوا مشتركين في شبكة الاتصالات ذاتها... إلخ (Bollier 2010; Graham 2012; Crampton et al 2012). فالتقنية المستخدمة والمعايير المطبقة تؤثر على طبيعة البيانات. فعلى سبيل المثال، يمكن أن يتأثر تحديد المشاركات الأكثر قراءة أو مشاركةً على وسائل التواصل الاجتماعي بشكل كبير من خلال خوارزميات الترتيب وليس الاهتمامات البسيطة (Baym 2013). وعلى النحو ذاته، فإن واجهات برمجة التطبيقات تنظم ماهية البيانات التي يتم استخلاصها، ففي تويتر، على سبيل المثال، يتم التقاط الوسم المرتبط بالحدث وليس جميع التغريدات ذات الصلة (Burns 2013)، وأيضًا فقد بين (Gonzales – Bailon et al 2012) أن اختلاف طرق الوصول إلى بيانات تويتر - البحث في واجهات برمجة التطبيقات مقابل تدفق واجهات برمجة التطبيقات - يؤدي إلى الحصول على مجموعات مختلفة من النتائج. ونتيجة لذلك، لا توجد أي ضمانات بأن فريقين من الباحثين، على سبيل المثال، الذين يحاولون جمع البيانات ذاتها في الوقت ذاته سوف يكونون قادرين في نهاية المطاف على الحصول على مجموعات البيانات ذاتها (Burns 2013). وعلاوةً على ذلك، فإن اختيار البيانات الوصفية والمتغيرات التي يتم توليدها والبيانات الوصفية والمتغيرات التي يتم تجاهلها يؤدي إلى رسم صورة معينة. ومثل هذه التحيزات تؤدي إلى انحراف أو حتى إلى نشوء أفكار واتجاهات وأشخاص ونماذج مستترة لم يتم تصويرها أو تمثيلها في مجموعة البيانات (Graham 2012).

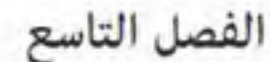
وفيما يتعلق بمصداقية البيانات، فإن الحال لا يتمثل دوماً في أن البيانات تعبر بصدق عما تم استخدام البيانات من أجل قياسه. فعلى سبيل المثال، هناك علامات استفهام



حول مدى حقيقة ودقة تعبير المشــاركات في وســائل التواصل الاجتماعي عن وجهات نظر الأشــخاص، وعلى مدى الوثوق في مثل هذه المشــاركات. وقد حذر Manovich (2011:6) قائلاً إن "المشــاركات، والتغريدات، والصور التي تم تحميلها، والتعليقات، وجميع الأشكال المختلفة من المشاركات على الإنترنت لا تعتبر نوافذ شفافة بذاتها، وبدلاً من ذلك، فهي غالباً مرتبة بعناية وتتم إدارتها بشكل منتظم وذلك من خلال الانتقادات نفسها التي تنطبق على بيانات المقابــلات ومجموعات التركيز" (Focus Group). وأبدى آخرون شــكوكهم حيال صحة بيانــات علم المواطنــة (Citizen Science) والتعهيد الجماعــي، وحذروا من جهل وعدم مقدرة الحشــود على التنبــؤ ومخاطر التفاوت (فيما يتعلــق بالتغطية)، والمتغيرات (فيما يتعلق بالجودة والتجانس)، وتوليد البيانات المتحيزة المتأصلة في مشروعات المســاواة الكــبرى، التي تفتقر إلى القادة وتتشــكل من مجموعة ضيقة من الســكان ذوي التباين في المهارات والدوافع (Carr 2007). وفي الواقع، فإن وسائل التواصل الاجتماعي يهيمن عليها نخــب رقمية غير ممثلة (Crutcher and Zook 2009) في حين أن هناك مجموعات كبيرة من المجتمع مثل الفقراء وكبار السن لا ينخرطون في وسائل التواصل الاجتماعي أو يفتقرون إلى الهواتــف الذكية، ومن ثم فإن أي بيانــات يتم جمعها لا تعكس بالضرورة وجهات نظر هؤلاء الأشــخاص أو أحيائهم (Crawford 2013). ومن ثــم فمن الضروري للغاية أن يتم تحديــد ووضع مثل هذه التحيزات في الاعتبار وذلك عند تحليل البيانات من أجل كســب فهم أكثر شمولية للظاهرة، وخصوصاً إذا ما كان سيتم استخدام مثل هذه البيانات في أعمال السياسة العامة والقرارات التي تؤثر على حياة الناس كالنتائج الصحية.

وعلاوة على ذلك، يوجد بعض المخاوف بشأن جودة وتجانس المحتوى والبيانات الوصفية التي تم تكوينها عبر أفراد متنوعي المهارات والدوافع، ومخاوف بشأن كيفية توفير درجات موثقة من الاعتمادية وتكوين إحســاس بالمصداقية (Dodge and Kitchin 2013). وقد أدى هــذا بالبعض إلى افتراض أن عمل التعهيد الجماعي "الهواة" ســيكون إجراؤه بشــكل أفضــل إذا بُذل عــلى تصحيح وتنويع البيانات وليس على تكويــن البيانات (Carr 2007). فالمثــال الــذي ألقــى Carr الضوء عليه هو موســوعة ويكيبيديا، إذ عــلى الرغم من كونها شهيرة وشــاملة، إلا أنها تتوسع بشــكل غير مدروس يُماثل الاهتمامات المختارة للمشاركين، ويوجد بها مقالات غير كاملة وأحياناً مكتوبة بلغة ركيكة وتافهة ومطعون في صحتها بشكل





كبـير، مما يقـوض حجيتها وإمكانية اسـتخدامها أو نفعها. وادعـى Carr أنه " إذا كانت موسـوعة ويكبيديا غير مجانية، فمن غير المحتمل أن قراءها سـيكونون أكثر تسامحاً تجاه تلك العيوب." (2007:4). ويمكـن أن تُعاني خريطة الطريق المفتوح (OpenStreetMap) سـوءَ التغطية في بعض الأماكن حيث يوجد عدد قليـل من المتطوعين. ويوجد أيضًا بعض المخاوف بشـأن اسـتدامة أعمال التعهيد الجماعي التطوعية، حيث ادعى Carr 2007 أن الارتباطات التي تربط مجموعة افتراضية من الناس سوياً عادةّ ما تكون ارتباطات سطحية، وهـي تفتقر إلى العمـق والالتزام الإجباري، وعادةّ ما تكون عرضةّ للتشـتت، وهي تعتمد بشـكل كبير على مجموعة أساسية صغيرة لكي تحافظ على سير المشروع وتوفر الجزء الأكبر مـن العمل. وعلى النقيض من ذلك، لاحـظ البعض أنه فيما يتعلق بـ OpenStreetMap، فـإن جودة البيانـات المقدمة تطابق جودة البيانات التي تقدمهـا الشركات الاحترافية وأن التغطية متنوعة (Haklay 2010, Mooney at al 2011).

فهذه المناقشـة تلقـي الضوء على أن كون مجموعة البيانات كبـيرة للغاية من ناحية الحجم، فإن هذا لا يعني بالضرورة أنها عشـوائية أو ممثلة للضاهرة أو صحيحة أو ذات مصداقية أو موثوقة. وكما لاحظ (2012) boyd and Crawford: "كون البيانات الكبيرة تزودنـا بكميات هائلـة من البيانات، فإن هذا لا يعنـي أن القضايا المنهجية لم تعد ذات صلة. ففهم العينة على سـبيل المثال، يُعد أكثر أهمية الآن عما كان عليه الأمر في الماضي" (ص 668). وبالنسبة لـ (2009: 615) .Kelling et al فإن هذه القضية تعتبر قضية أصل ومَنْشَأ أي "المعلومات المتعلقة بأصل وتعريف وملكية وبنية" مجموعة البيانات التي تمكن محللي هذه البيانات مـن معرفة كيفية تكوين هذه البيانات ومعالجتها، وتمكنهم كذلك مـن تقييم جودة وصحة البيانات. وعلى الرغم مـن ذلك، فإن مصدر معظم مجموعات بيانات البيانات الكبيرة مقيد بهؤلاء الذين قاموا بإنشاء البيانات، مما يجعل من الصعوبة بمكان تقييم صحة الدراسات المنشورة التي اعتمدت على هذه البيانات. فأحد التحديات الرئيسـية التي تواجه البيانات الكبيرة يتمثل في المقدرة على معرفة تفاصيل سُلالة الأصل والذي يشمل تقييمات مستويات الأخطاء والشكوك في مجموعة البيانات، وإنشاء وسائل فعالة للتحقق من وتنظيف ومعايرة البيانات عالية السرعة وشديدة التنوع.



دمج وتوافقية البيانات (Data Integration and Interoperability):

إن أحد الوعود التي انبثقت عن ثورة البيانات هو إنشاء مجموعات بيانات ذات ترابط قوي مما يمكن هذه المجموعات من أن تتحد سوياً من أجل إنشاء قيم ومعارف إضافية.

وكما اتضح من خلال تطور البنى التحتية للبيانات ومدمجي البيانات، فإن العديد من البيانات عرضةٌ لكي تندمج مع بعضها، واتضح أيضًا أن الكثير من البيانات تظل منعزلة في مستودعات أو يمكن دمجها فقط من خلال معالجة إضافية كبيرة. ومن أجل أن يتم دمج البيانات في مجموعات بيانات جديدة، أو استخدام هذه البيانات بالاشتراك مع أي بيانات أخرى، فإن هذه البيانات تتطلب حقول تأشيريه مشتركة وتوفر معايير البيانات، وبيانات وصفية متسقة، وتوافق على مستوى التنظيم والصيغة. وفي حين أن تقدماً كبيراً قد حدث في إدارة البيانات وتطبيق معايير البيانات التي قادتها منظمات مثل المنظمة الدولية للمعايير (ISO) ورابطة الشبكة العالمية (W3C) واتحاد المعلومات المكانية المفتوحة (OGC) في سبيل المحاذاة بين البيانات والبيانات الوصفية وتحسين التوافق بين المنصات التقنية المختلفة التي تعالج البيانات وتحتفظ بها، وتطوير البيانات المترابطة، إلا أن الكثير من مجموعات البيانات لا يمكن ربط بعضها ببعض. فهي تشكل ما أطلق عليه Singh (2012) مصطلح "البيانات الجانحة" (Stranded Data).

إن معظم دراسات البيانات الصغيرة لا تزال تُنشئ بيانات جانحة كونها تستخدم لغات توصيف خاصة بها وأنظمة تنظيم البيانات التي تناسب الأغراض التي سيستخدمونها فيها، ولكن القليل من هذه الدراسات تضع في اعتبارها قضايا التوسع والربط. وهذه هي أيضًا الحال غالباً داخل المنظمات التي يوجد بها إدارات منفصلة تستخدم أنظمتها الإدارية والتصنيفية الخاصة. فعلى سبيل المثال، فإن كل واحدة من هيئات التخطيط التي يبلغ عددها ثمان وثمانين هيئة في إيرلندا تستخدم نظامها الخاص لاستخدام الأراضي وتصنيف المناطق، وعلاوة على ذلك، فهذه الهيئات توظف العديد من المنصات التقنية المختلفة لغرض إدارة وتخزين وتخطيط البيانات، فربط جميع هذه البيانات وتوحيدها في نظام وطني واحد ليس مهمة سهلة على الإطلاق. وبالمثل، فإن توسيع مجموعات البيانات الوطنية سوف يواجه مشاكل مماثلة. فعلى سبيل المثال، فإن ربط الإحصائيات السكانية



في إيرلندا وإيرلندا الشمالية سوياً من أجل إنشاء مجموعة بيانات واحدة للجزيرة بالكامل يفرض عدداً من التحديات الفنية وذلك يرجع إلى أن ٣٠٪ من الأسئلة متطابقة، و٣٥٪ أخرى من الأسئلة يمكن أن يتم مضاهاتها وذلك من خلال تقسيم خطط التصنيف وإعادة تجميعها في الخطط المشتركة، وحتى إذا تم هذا، فإن البيانات ستكون ذات جغرافيات إحصائية ضمنية مختلفة يتم تجميعها في وحدات مختلفة الأحجام (Kitchin et al 2007).

وتتضاعف هذه القضايا عبر قارة أوروبا، حيث تتبع كل دولة وكل مؤسسة إجراءات مختلفة (وحدات وتصنيفات)، وتعتمد فترات زمنية متفاوتة (عدد المرات التي يتم فيها إنشاء البيانات)، وتستند إلى أماكن متباعدة (الجغرافيا الإحصائية)، وتستعمل تقنيات مختلفة (الصيغ). ومن ثم فلقد استنتج Rubert (2012: 118) أنه حتى مع وجود بيانات حكومية شاملة، إلا أنها في غالب الأحيان تتصارع تحت وطأة الأعداد التي لا تحصى من مجموعات البيانات المتضاربة وغير المتوافقة وغير الصالحة للمقارنة. وتسعى بعض المبادرات مثل توجيهات إنسبير (INSPIRE) (البنية التحتية للمعلومات المكانية في المجتمع الأوروبي) إلى تكوين بيانات مشتركة عبر الجغرافيا الأوروبية الإحصائية المشتركة، ولكن الوصول إلى توافق في هذا الشأن يُعتبر عملية بطيئة وذلك يرجع إلى أن جميع الولايات لا ترغب في الانتقال بشكل أحادي (أي من طرف واحد) إلى نظام جديد للبيانات وخلال العملية يوقفون سجلاتهم الزمنية ويفقدون مقدرتهم على إجراء تحليلات زمنية متتالية، إضافة إلى أن كل ولاية لها أولويات مختلفة فيما يتعلق بالبيانات التي ترغب في إنشائها وتتبعها.

وتوجد هذه التحديات نفسها فيما يتعلق بالبيانات الكبيرة، والتي ربما تكون غير متجانسة وغير منظمة بشكل كبير للغاية، ومتفاوتة فيما يتعلق بالجودة، وهي أيضًا مرنة وقابلة للتوسع. فربط وتجميع مثل هذه البيانات ليس مهمة سهلة على الإطلاق وهي تتطلب تطوير مجموعة جديدة من الأدوات والأساليب، وكذلك إنشاء وتبني معايير جديدة للبيانات. فأحد الحلول المطروحة، والذي أشار إليه Short et al (2011) هو دمج البيانات المخصص والذي يمكن المستخدمين من التحكم بسرعة وبشكل مباشر في أي من البيانات التي سيتم دمجها، وذلك على الرغم من أن هذا يتطلب أن تكون مجموعة البيانات تتمتع بسمات معينة مثل البنية التحتية الدلالية المؤثرة. إن ربط وتجميع مثل هذه البيانات من الممكن أن يؤدي إلى التلوث المتقاطع (Cross-Contamination) بسبب البيانات



ذات الجـودة الضعيفة والحالة المتهالكة مما يؤدي إلى فسـاد وتلويث مجموعات البيانات الفعالـة ويحولها إلى مجموعات بيانات ضعيفة. ومع ازديـاد طوفان البيانات، فإن إيجاد الحلول التي تسـاعد على دمج البيانات وتوحيدها وتحافظ على سلامتها وفعاليتها سيمثل قضيـة ملحة، وذلك على الرغم من أن عدم المقدرة عـلى ربط البيانات له جوانب إيجابية أيضًا وذلك فيما يتعلق بتقييد مراقبة البيانات وتحسين أمن البيانات (انظر الفصل العاشر).

التحليل الركيك والمغالطات البيئية (Poor Analysis and Ecological Fallacies):

جميع النتائج التي يتم التوصل إليها من الدراسات تكون دقيقة وسليمة وقوية بحسب البيانات التي تم استخدامها والتحليلات التي أُجرِيَتْ، وإذا ما وضعنا جودة وصحة البيانات جانبـاً، فخـلال تصميم أي بحث يكون مـن المهم للغاية اسـتخدام منهجية ملائمة وآليات مناسـبة. فالإخفاق في فعل ذلك سـيؤدي إلى مغالطات بيئية خطيرة عند تفسـير النتائج، بمعنى اسـتخلاص النتائج المضللة والتي في الواقع لا تدعمها البيانات المسـتخدمة. وهناك تاريـخ طويل من المسـاجلات التي تتعلق بمكامن القوة والضعف النسـبية في العديد من الأنماط والأسـاليب المنهجية المختلفة، والظروف التي يمكـن فيها تطبيق مثل هذه الأنماط والأسـاليب بشكل صحيح. وعلى وجه العموم، هناك الآن إرشادات وقواعد محددة بشكل واضح تتعلق بكيفية تكوين وتحليل أنواع مختلفة من البيانات عبر أنظمة متعددة. وعلى الرغـم من ذلك، وفيما يتعلـق بالبيانات الكبيرة، فـإن علوم البيانات لا تـزال في أطوارها التمهيدية، ومثل هذه المعايير لا تزال في طور التشـكيل والصياغة، فآليات جديدة يتم الآن اختراعها، وتوسيعها، وتقييمها.

وهناك أربع قضايا منهجية تتعلق بالبيانات الكبيرة جذبت الانتباه، فأولى هذه القضايا تتمثل في الإجابات المتفاوتة بشـكل عميق والتي تؤدي إليها الافتراضات والمقاربات داخل النماذج (Silver 2012). وثانياً، مدى قدرة الآليات والأسـاليب المسـتخدمة على إحداث الاستسقاط، أي رؤية واستخلاص أنماط من البيانات العشوائية التي لا معنى لها، أو إحداث فرط التعلـم أو الملاءمة (Overfitting) والتي تحـدث عندما يصف نموذج إحصائي خطأ عشـوائيًا أو ضجيجًا بدلاً من العلاقة الكامنـة (boyd and Crawford 2012)، أو الخطأ في التعامل مع التشـويش على أنه معلومات (Silver 2012; Siegel 2013). وثالث هذه

القضايــا يتمثــل في الاتهامات بأن بعــض الآليات المعينة تؤدي إلى ما هــو أكثر بقليل من تجريف البيانات (Data Dredging) (Webster 2011). وأخيراً، إذا كان بإمكان الآليات أن تتوســع خــارج نطاق بيانات ضبــط الاختبار إلى مليارات الســجلات المتدفقة في العالم الواقعي (Bryant et al, 2008).

إن النــماذج تهدف إلى توضيح والتعبير عن العلاقــة بين المتغيرات وكيفية عمل النظام. فكل نموذج يتم تأسيســه عــلى نظرية معينة، وبناؤه باســتخدام مجموعة من الافتراضات والمقاربــات التي تبين كيفيــة تفاعل المتغيرات، في حين أنه يتــم اختبار صحة النموذج من خلال البحوث التجريبية، فإذا كانت النظرية والافتراضات غير صحيحة، فإن جميع التنبؤات ستكون غير صحيحة في هذه الحالة كذلك، وحتى التغييرات الصغيرة للغاية على الافتراضات في نمــوذج يمكن ما أن تؤدي إلى نتائــج مختلفة تماماً. ففي حالة الأزمة المالية العالمية التي حدثــت في العام 2008، ادعــى (Silver 2012) أن أحد العوامــل المهمة للغاية في إحداث مثل هذه الأزمة كان اســتخدام نماذج تمويل دمجت المخاطر بحيث يمكن حساب احتمالية الدخــل المحتمل مــع الارتياب لأن المخاطر المحتملة من الصعــب تقييمها. ونتيجةّ لذلك، اعتمدت النماذج على افتراضات غير يقينية ولكن تم معاملتها كما لو كانت مخاطر معروفة عــلى وجه اليقــين. والنتيجة كانت افتراض أن الســندات المالية الجديــدة التي يحيط بها ارتياب نظامي عالٍ للغاية هي أصول ذات مخاطر منخفضة، وهذا افتراض ثبت أنه خاطئ. وفي الأنظمة المعقدة، فإن مثل هذه الأخطاء لا يتم قياســها على أســاس درجات ولكن يتم قياسها على أساس ترتيبها حسب الأهمية. وقد قدرت مؤسسات ستاندرد أند بورز وموديز (S&P and Moody's) مخاطر التخلف عــن الدفع والمرتبطة بالتزامات الديون المضمونة (CDOs) بأقل من قيمتها بعامل من مائتين (Silver 2012: 45). وبغض النظر عن جودة البيانــات الضمنية، فإذا كان النموذج خاطئًا، فأي نتائج يتم اســتخلاصها من هذا النموذج ستكون خاطئة كذلك.

لقد افترض Granville (2013) and Taleb (2013) أن لعنة أو مأساة البيانات الكبيرة مُضاعفــة، نتيجة العلاقات الإيجابية الخاطئة بين البيانات؛ وأنماط مجموعات البيانات التي تكــون عرضية بشــكل تــام، وليس لها أي قوى تنبؤيــة، ولا يمكن تكرارهــا، وربما يمكن أن



تكــون قناعاً لنماذج أضعف ذات أهمية، فمجموعات البيانات تشــتمل غالباً على البيانات التــي يبدو أنها مرتبطة بعضهــا البعض، حتى ولو كانت العلاقة بينها عشــوائية في الواقع. ومــع زيادة حجم مجموعة البيانات، تتضاعف الأخطاء الإيجابية، إلى الحد الذي تكون فيه أي علاقــة تقريباً بين البيانات مهمة من الناحيــة الإحصائية وفق معايير اختبارات الارتباط التقليدية (Pentland 2012; Rajaraman et al. 2012; Taleb 2013). وعلى النحو ذاته، ففي أســاليب التنقيب عن البيانات، مثل شجرة القرارات (تسلسل اتخاذ القرار)، يمكن أن يتسم النموذج بفرط الملاءمة كي يتم استنباط نتائج استثنائية تحول إلى نتائج قابلة للتعميم مما يؤدي هذا إلى حدوث مغالطات بيئية. فالبيانات الكبيرة في هذه الحالة لا تعني معرفة أفضل، ولكنها تعني بالأحرى ارتباكاً أكثر وذلك من جراء صعوبة تحديد دالة حقيقية وسط ركام من الدلالات. ولذا فإنه: " أحياناً يكون الأقل هو الأقوى تأثيراً" (Verhulst تم الاقتباس من Bollier 2012:14).

إن أحد الجوانب المهمة في مشــكلة الأخطاء الإيجابيــة ليس حجم مجموعة البيانات فحسب، بل كيفية معالجة وتحليل مجموعة البيانات هذه. فأحد أجزاء إستراتيجية الكثير من تحليلات البيانات الكبيرة يتمثــل في فحص مجموعة البيانات قبل صياغة الافتراضات (انظر الفصل الســادس)، ومن الطــرق المتبعة للقيام بهذا الأمر حســاب الارتباطات بين جميع المتغيرات وذلك من أجل تحديد العلاقات المحتملة بين البيانات الجديرة بالدراسة الإضافية. وتشــكل هذه العملية ما يُعرف بتجريف البيانات (Data Dredging) والتي تسمى أحياناً تصيد البيانات، وهي ممارسة لا يحبذها بشدة الكثير من الخبراء الإحصائيين (Webster 2011; Piatetsky-Shapiro 2012). فالمشــكلة التــي ترتبــط بهــذا المنحى مشــكلة مزدوجة. فمن جهة، هــي تحدد آلاف الأخطاء الإيجابيــة المحتملة وذلك دون تحديد أيٍّ من هذه الأخطاء عشوائي وأيّ منها ذو مغزى. ومن جهة أخرى، يمكن أن تؤدي إلى التنقل من متجر لآخر طلباً للأرخص وعرض النتائج التي تؤكد اعتقاد المحلل، مع وجود دليل من التحليل المستخدم من أجل أن يجعل الادعاء يبدو معقولاً ومدروساً (Webster 2011; Taleb 2013). فالســؤال الشائك الذي يطرح نفســه هنا، كما أشار (Granville 2013)، هو كيف يمكن التمييز بين دالة حقيقية ودالة عرضية وسط الكميات الهائلة من البيانات؟ وكيف يمكن ألا ينقاد المرء للاستسقاط وفرط التعلم؟



إن هنــاك بعض الحلول الممكنة. حيث اقترح (2015) .Rajaraman et al اســتخدام مبدأ بونفيروني (Bonferroni)، والذي يســعى إلى تحديد الأحداث النادرة جداً لدرجة أنها لا يمكــن أن تحــدث في البيانات العشــوائية، وذلك من أجل فرز الارتباطــات الفعلية عن الارتباطات العشــوائية. كما كان (2013) Lusk من أنصار اســتخدام مســتويات معيارية معززة، وذكر نقطة انطلاق معيار خمسة سيجما (Five Sigma) المستخدم حالياً في الفيزياء مــن أجل مواجهة الارتباطات الوهمية وإعلان الاكتشــافات (احتــمال كون النتائج خاطئة وناتجــة عن التغــيرات الإحصائية فقط هــو واحد في كل ثلاثة ملايين وخمســمئة ألف أو 1:٣,٥٠٠,٠٠٠). ومن الحلــول المطروحة الأخرى هو التحقق الإضافي من خلال مجموعات بيانات أخرى، أو التقســيم العشــوائي لمجموعة البيانات الحاليــة إلى عينات ومعرفة إذا ما كانــت العلاقة أو الارتباط ســيظل موجوداً عــبر العينــات (Jensen 2000). وفيما يتعلق بفرط التعلم، ناقش (2010) Miller إســتراتيجية استخدام الخلفية المعرفية، والتي تتكون من اســتخدام الحقائق المعروفة عن مجموعة بيانات مُســتقاة من الافتراضات عن النظام، والحقائــق الملاحظة، ومعرفة الخبير، أو النظريــة، أو المقاييس الممتعة والمرغوب فيها التي تقيم البســاطة والتيقن والفائدة والحداثة للأنماط المستحدثة، وذلك من أجل فصل الأنماط الهادفــة عن التشــويش. وعلى الرغم من ذلك، فمن الواضــح أن هناك حاجة لبذل أعمال إضافية من أجل تطوير أساليب لتقييم المغزى والدلالة في عصر البيانات الكبيرة.

وإذا وضعنــا في الاعتبــار أن تحليلات البيانات لا تزال في مراحلها الأولى من التطور، وتم تطبيقهــا بالدرجــة الأولى على بيانات الضبــط والاختبار، يظل أيضًا هنــاك بعض المخاوف بشــأن فعالية مثل هذه الأساليب في مجاراة البيانات الفوضوية وغير المنضبطة. وكما أشار (2008: 4) .Byrant et al، فإن "الكثير من الخوارزميات لا تتوسع خارج نطاق مجموعات البيانــات التي تتكون مــن بضعة ملايين من العنــاصر، أو لا يمكنها أن تتحمل التشــويش والثغــرات الإحصائية الموجودة في بيانات العــالم الواقعي". فإذا كان هناك أي اختلالات أو تحيزات موجودة في الخوارزميات، فمن شأنها أن تؤدي إلى انحراف النتائج وتفسيرات هذه النتائــج (byoed and Crawford 2012). ويرتبط بهذا التخوف أيضًا القلق من أن هناك اعتــماداً مفرطاً والكثير جداً من الموثوقيــة الموضوعة في الخوارزميات وذلك من أجل إجراء التحليلات، واســتخراج المنافع والمزايا التي تجلبها الخــبرات والمعارف والفطرة البشرية من



أجـل فهم ظاهرة مـا (Lohr 2012). ومن ثم، هناك حاجة لإجـراء المزيد من العمل من أجل تأسـيس وتقييم صحة التحليلات الجديدة والثقة في النتائج التي يمكن استخلاصها من هذه التحليلات.

ومن الانتقادات الأخرى الموجهـة إلى تحليلات البيانات الكبيرة هو تركيز هذه البيانات عـلى الآنية (هنا والآن)، وذلك بالنظر إلى حداثتها، وعلى وجه العموم، فإن البيانات الكبيرة تفتقر إلى العمق التاريخي. فشبكات الاستشعار الجديدة أو بيانات التعهيد الجماعي يُمكن أن توفر بيانات مسـتمرة وثرية، ولكنها تتسم بقصر الفترات الزمنية التي تغطيها. ونتيجةً لذلـك، فإن الكثير من التحليلات تكون مناسـبة في وقتها ولكنهـا تفتقر إلى العمق الزمني. وهذا أمر لا غبار عليه ومستحسـن إذا كان الشـخص مهتـماً بالوقت الحاضر بتتبع حدث ممتد باستخدام وسائل التواصل الاجتماعي (على سبيل المثال حدث سياسي كبير مثل الربيع العربي)، أو أن يكون الشـخص قائماً بتحليل عملية مادية لا تتغير بمرور الوقت (مثل تجربة فيزيائية)، ولكن سـيكون الأمر مسبباً للمشـاكل إذا تم استخدامه من أجل فهم العمليات التي تمتد لآلاف السـنين أو أكثر (على سبيل المثال الأنظمة البيئية والجيولوجية). وينطبق هـذا الانتقاد أيضًا، بالطبع، على دراسـات البيانات الصغيرة، والتـي تعتبر أيضًا لقطات تم اتخاذها في أماكن معينة وأوقات محددة.

المهارات والموارد البشرية (Skills and Human Resourcing):

في حـين أن بعـض تحليلات البيانـات يتم تركهـا للخوارزميات، وخاصـةً العمل المتعمق للمعالجة والحسـابات، إلا أن أعمال التوجيه والتفسير لا تزال إلى حد كبير حكراً على المحللين البـشر. فبالاعتماد على مهاراتهـم وخبراتهم ومعارفهم، يتخذ الباحثـون والمحللون القرارات المتعلقـة بالأماكن المثلى التـي ينبغي التركيز عليها، وكيفية صياغة وإجـراء التحاليل، وإدراك المراد من النتائج والعمل على هذه النتائج. فالبشر يظلون هم الفاعلين الأساسـيين في تأسيس وتشغيل والمحافظة على المشروعات التي تقودها البيانات، فعلى سبيل المثال، يتطلب تأسيس بنية تحتية للبيانات مهندسي برامج وأجهزة ومسـئولين رقميين، وأخصائيين في السياسة وخبراء مجـالات علميـة ومحللين. فهؤلاء العاملـون يضمنون تصميماً جيداً للنظـام كي يُنتج بيانات مفيدة وصحيحة وموثوقة، ويوفرون الوسـائل الضرورية لتخزين ومشاركة البيانات، ويؤمنون



مجموعــة من الأدوات التحليلية والإدارية المناســبة، التي يمكنها اســتخلاص القيمة والمعرفة. وهم يعرفون أيضًا ماهية الأسئلة التي ينبغي توجيهها وما الذي يتعين فعله بالإجابات على تلك الأســئلة. ولذا، فإن المواهب البشرية بضاعة رائجة. ولكن وكما أشــارت بعض التقارير، هناك نقص متزايد في الموظفين الذين يتمتعون بمهارات في البيانات وخاصةً محللي ومديري البيانات الذين بإمكانهم تحويل البيانات والمعلومات إلى معارف وعلوم (Gantz and Reinsel 2011; Manyika et al. 2011; Shah et al. 2012; Rooney 2012).

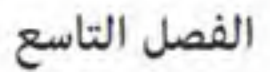
وقــد أكد Manyika et al. (2012:3) " أن الولايــات المتحدة وحدها تعاني نقصًا يبلغ ١٤٠,٠٠٠ إلى ١٩٠,٠٠٠ شخص يتمتعون بمهارات تحليلية متعمقة وكذلك ١,٥ مليون مدير ومحلل من أجل تحليل البيانات الكبيرة واتخاذ القرارات بناءً على اكتشافاتهم". واستطردوا بالإشــارة إلى أن مثل هذه المواهب "من الصعب إيجادها إذ تســتغرق ســنوات عديدة في التدريب أو إعادة التدريب" (ص ١٠). وقيم Shah et al. (2012) عدد ٥٠٠٠ موظف في اثنتي عشرة شركة عالمية فيما يتعلق بمقدرتهم على معالجة تحليلات البيانات بشــكل مثمر حيــث قاموا بوضع الموظفين في ثــلاث مجموعات. الأولى، التجريبيون غير المعترضين الذين يثقون في التحليلات أكثر من الأحكام الشــخصية (٤٣٪). والثانية، متخذو القرار المتعمقون الذين يتجاهلــون التحليلات ويتصرفون بما يمليه عليه حدســهم الفطري (١٩٪). والثالثة، المتشــككون المطلعون (٣٨٪) الذين يوازنون بين الأحكام الشــخصية والتحليلات، ويسعون إلى تفســير البيانات ووضعها في سياقها المناســب عند اتخاذ القرارات. وبشكل مقلق، فإن أولئك الذين يدعون البيانات تعبر عن نفسها وأولئك الذين يتجاهلون البيانات والتحليلات بشكل روتيني يفوقون في عددهم أولئك القادرين بشكل أفضل على فهم البيانات الكبيرة. وعلاوةً على ذلك، لاحظوا أن الشركات تعاني من أربعة مشــاكل تحول دون الاســتفادة من إمكانيات البيانات الكبيرة وهي:

- أولاً: أن المهــارات التحليلية تتركز في عدد محدود للغاية من الموظفين ولا تنتشر في باقي الموظفين.
- ثانياً: أن حلول تقنية المعلومات تتركز بشكل تقليدي على التقنية وبشكل أقل بكثير على المعلومات في حين أن المعلومات هي المجال الذي يتم من خلاله جني القيمة.

- ثالثـاً: أن تخزيــن البيانات يحدث كيفما اتفق ومــن الصعوبة بمكان على أي موظف أن يحدد مكان المادة إذا لم يكن هذا الموظف هو المسئول عن جمعها.
- رابعاً: أن المديرين يفهمون قيمة الموهبة ورأس المال والعلامة التجارية، ولكن لا يفهمون قيمــة المعلومــات، ولا يبذلون جهوداً كافية من أجل فهــم البيانات أو ترتيب هذه البيانات بشكل فعال وكفء.

ويوجــد مثل هــذا النقص في المهارات أيضًا في المؤسســات الأكاديميــة، حيث إن خبراء المجــالات المختلفة تدربوا في وقت ندرة البيانات وهم غير مســتعدين بشــكل كبير لعصر البنى التحتية للبيانات والبيانات الكبيرة وذلك باستثناء عدد لا يتجاوز أصابع اليد الواحدة من العلماء والمراكز.

إن القلــق الذي ينتاب الكثير من المعلقين هو أن المنافع المحتملة من الشركات والعلوم التي تقودها البيانات لن يتم تحقيقها بشــكل كامل بسبب نقص المواهب البشرية وخاصةً علــماء البيانات الذين تجتمع لديهم مهارات مبرمجي البرامج والإحصائيين والرواة المحترفين وذلك من أجل اســتخلاص شــذرات الذهــب المختبئة تحت جبالٍ مــن البيانات (Cukier 2010)، والمديرين الذين يفهمون كيفية تحويل هذه الشــذرات إلى قرارات حكيمة. وفيما يتعلق بالمديرين، وكما أشــار (2012: 23) .Shah et al، فإن "الاستثمار في التحليلات يمكن أن يكــون بــلا جدوى، ويمكن حتى أن يكون ضــاراً، إلا إذا كان بإمكان الموظفين دمج هذه البيانــات في العمليــة المعقدة لاتخاذ القــرارات". فالجامعات بدأت الآن في إنشــاء برامج جديدة ومراكز بحثية لعلوم البيانات، ويســعون إلى تعديل الدورات التدريبية الحالية لكي تشــتمل على التدريب على مجموعة المهارات الجديدة هذه، وذلك في مســعى إلى تنقيح بعــض الثغرات المتعلقة بالمهــارات. ويتم تنظيم بعض هذه البرامج بالمشــاركة مع بعض الشركات. فعلى سبيل المثال، بدأت شركة آي بي إم (IBM) بالفعل في العمل مع العديد من الجامعات من أجل تطويــر دورات ومقررات تعليمية، وهناك بعض الشركات الأخرى مثل شركــة تيراداتا (Teradata) التي تقدم برامج تدريبية معتمدة من الجامعات (Bertolucci 2013)، وذلــك عــلى الرغم من أن هــذه البرامج من الممكن أن تــؤدي إلى مخاطرة توجيه الطلاب تجاه رؤى وبرامج خاصة ببعض الشركات.



الجدول رقم (١ - ٩)

الخبرات اللازمة لبناء البنى التحتية للبيانات وإجراء بحوث البيانات الكبيرة

خبرات المجال Domain Expertise	• معرفة نظرية وعملية عميقة بالمجال (المجالات) ذات الصلة. • المعرفة التامة بأنواع البيانات التي ستتم دراستها وأصل هذه البيانات وأهميتها للمجال (المجالات) ذات الصلة. • المقدرة على تحديد الفجوات المعرفية. • المعرفة التامة بالأدبيات والاتفاقيات النظامية. • المقـدرة على تعليم الآخرين الذين يملكـون خلفيات متعددة من أجل استيعاب كل ما ذكر أعلاه.
خبرات البيانات Data Expertise	• فهـم كيف يتم تجميع البيانات والمحافظة عليها، وكذلك فهم العلاقات بين الأشياء المادية والتمثيل الرقمي لهذه الأشياء (إذا كان ذلك ممكناً). • المعرفة التامة بنماذج البيانات و/ أو اتفاقيات وصف البيانات. • فهم كيف يتم الوصول إلى البيانات ذات الصلة وتخزين هذه البيانات. • المقدرة على تسـهيل مشـاركة البيانات وتصحيح الأخطاء بشكل يدوي، سواءً كان هذا أثناء أم بعد المشروع. • المقدرة على التنبؤ بالاستخدامات المستقبلية أو البديلة للبيانات. • فهم الأشكال الجديدة من المطبوعات التي يمكنها أن تحتوي على البيانات.
الخبرات التحليلية Analytical Expertise	• فهم نقاط القوة والضعف في أدوات البحث الفردية. • اختيار وتعديل الأدوات المناسبة لدعم الأهداف البحثية. • التنبـؤ بالمشـاكل التي يمكنها أن تنشـأ مع اسـتخدام الأدوات المختارة لتنفيذ مهام المشروع. • توقع واكتشـاف معـدلات الأخطاء في خوارزميـات البيانات وتحليلات البيانات، واختيار الطـرق الإحصائية التي تفسر هذه الأخطاء كلما كان ذلك مناسباً. • تعليم الآخرين كيفية تفسير نتائج البحث.



خبرات إدارة المشاريع Project Management expertise	• المقدرة على صياغة معالم المشروع. • المقــدرة عــلى تحديد الأهــداف المناســبة والمواعيد النهائية وتنســيق مسارات العمل المتوازية إذا كان ذلك ضرورياً. • المقــدرة على اختيار أفضل الإســتراتيجيات المناســبة للاتصال والتوثيق للمشروع. • إجادة الأدوات البحثية التعاونية. • الإرادة القوية للعمــل من أجل تحقيق النتائج التي تفيد جميع أعضاء الفريق.

المصدر: مقتبس من (20 – 16 :Williford and Henry (2012.

إن علــم البيانــات ليس هو الخبرة الوحيدة التي يتعين تأسيســها. وبحســب ما ادعى (Williford and Henry (2012، فإن تأســيس بنى تحتيــة للبيانات وإجراء البحوث على البيانــات الكبــيرة يتطلبان دمج أربعة أنــواع من الخبرات من أجل تحقيــق النجاح (انظر الجــدول ١-٩). وهذه الخبرات الأربعــة كما ادعى (Rubert (2013 يمكن أن يتم تحقيقها فقط من خلال منهج متعدد التخصصات والقطاعات، مما يســمح بتوحيد المهارات والرؤى التــي يمكنها أن تتقاطع مع المجالات وتحقق التعاون بين المؤسســات الأكاديمية والصناعية والحكومية. وحتى يأتي مثل هذا الوقت، أي عندما يتطور وينضج جيل جديد من المواهب المناسبة، وتتطور مناهج متعددة المجالات والأنظمة، فإن المشروعات ستظل عرضة للإخفاق أو سوء الأداء.

الخاتمة:

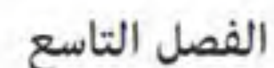
كشــف هذا الفصل عن أوجه النقــص الفنية في البيانات الصغيرة المتوســعة والبيانات الكبيرة. وناقش أن البيانات الكبيرة تحيط بها العديد من القضايا التي تحول دون الاستفادة منهــا وتطويرها في الوقــت الراهن، ومن هذه القضايا على ســبيل الذكر لا الحصر النطاق الضيــق إلى حد ما لمجموعات البيانات المتوافرة، والوصــول المحدود إلى البيانات، والجودة الضعيفة للبيانات، وصحة وموثوقية البيانات، وسُــلالة أصل البيانات المحددة بشكل سيئ،



والتبادلية الضعيفة، والتحليلات المشكوك فيها، والمغالطات البيئية، وقاعدة المهارات المطورة بشكل ضعيف، والاستجابة المؤسسية. وقد أدى ذلك ب (2012) .Dembosky et al للادعاء بأن "التحركات الأولية ربما تكون واعدة بشكل مفرط ولكن أداءها ربما يكون أقل من المتوقع"، وخصوصاً عندما يصبح هؤلاء الذين يتبنون حلول البيانات الكبيرة "مُتخمين بسبب التقنية التي لا تزال حديثة للغاية وغير مُنقحة، فالتقنية أضحت مرهقة ومثبطة أكثر من كونها مفيدة". وبمرور الوقت، فإن الكثير من هذه القضايا سيتم التعامل معها إلى حد ما وذلك من خلال الحلول الإدارية والفنية، ولكن ستظل هذه القضايا تشكل مخاوف وتتطلب اهتماماً مستمراً وعملاً إصلاحياً متجدداً. وينطبق الأمر ذاته أيضًا على مشروعات البنى التحتية للبيانات ومشروعات البيانات المفتوحة. وفي الواقع، وكما أشرنا إلى ذلك في الفصل الثالث، فإن الكثير من هذه القضايا التي تؤثر على البيانات الكبيرة تؤثر بشدة في مشروعات البيانات المفتوحة. فعلى سبيل المثال، فإن بعض مواقع البيانات المفتوحة لا تعد أكثر من كونها مقالب نفايات للبيانات ذات المتغيرات الانتقائية، إذ إن هذه المواقع منظمة بشكل سيئ للغاية وهي ذات بنية ضعيفة تماماً وبياناتها غير نظيفة، وهي تفتقر إلى مجموعات التسجيل الكاملة، والتوثيق، والسياسات المتعلقة بالحفظ، والنسخ الاحتياطي، والتدقيق، وإعادة الاستخدام، والخصوصية والسياسات الأخلاقية.

فنحن يتعين علينا أن نكون حريصين للغاية من أجل أن ندرك أن ثورة البيانات لا تزال في أطوارها التمهيدية وهي تتمتع بخصائص الجنين فيما يتعلق بالنضج والتطور، وفي الوقت الراهن، يبدو أن التسابق يتجه نحو طرح نظم تهدف إلى استعراض الإمكانيات وتأمين حصة سوقية مبكرة دون بذل الجهد الكافي لتفحص المصداقية والفعالية والمنفعة، أو لتفحص القضايا الإدارية والموارد البشرية التي تتعلق بالجاهزية المؤسسية ومجموعات المهارات المناسبة. وفي الواقع، فإن شعار بعض المشروعات هو أن تتجه نحو ما هو مستهجن بشكل مبكر ثم تتعامل مع القضايا التي تنشأ على أساس انتشارها.

ونتيجةً لذلك، فإن البحوث المركزة بشكل كبير والمصممة من أجل تحديد طبيعة ومدى أوجه القصور الفنية والمؤسسية للبيانات الكبيرة والبيانات المفتوحة والبنى التحتية للبيانات لا تزال متأخرة بشكل واضح مقارنة بأعمال التطوير الأخرى. ولكن مثل هذه البحوث





تظل حيوية للغاية لإنشـاء أنظمة تقدم اكتشـافات ونتائج صحيحة، ومن شـأنها أيضًا أن تضيف قيمة كافية للمؤسسـات التي تتبنى وتستخدم هذه البحوث. ويتعين أن يتم إجراء هـذه البحوث عـبر قطاعات متعددة وعـبر الأنواع المختلفة من البيانـات وذلك من أجل تأسيس القضايا والحلول المتعلقة بالبيانات والأنظمة المختلفة. وعلاوة على ذلك، يتعين أن يتم تخصيص اسـتثمارات من أجل زيادة وصقل مهـارات الموظفين الحاليين وتدريب جيل جديـد من علماء البيانات، ودون إجراء مثل هذه البحوث، فإن البيانات الكبيرة، والبيانات المفتوحة والبنى التحتية للبيانات ربما سـتعاني أضرار السمعة التي ستؤدي في نهاية المطاف إلى بـطء في نشرها وتبنيهـا. وهناك مجموعة أخرى من القضايا التي سـتؤدي الدور ذاته وهي القضايا المتعلقة بالآثار الأخلاقية والاجتماعية والسياسـية لأنظمة البيانات الجديدة، وهو ما يركز عليه الفصل التالي.


المكتبة الاقتصادية
ECONLIBRARY
قناة المكتبة على التيليجرام


المكتبة الاقتصادية



الفصل العاشر

المخاوف الأخلاقية، والسياسية، والاجتماعية، والقانونية

(Ethical, Political, Social, and Legal Concerns)

كما نوقش في الفصل الســابع، يتم تكوين وتوظيــف البيانات للعديد من الغايات التي تشمل حكم المجتمعات، وإدارة المنظمات، والتوسع في الأرباح، وتنظيم الأماكن. وفي جميع هذه الحالات، تُعَدُّ البيانات مدخلات أساسية في الأنظمة التي يتم تنفيذها، للمفارقة، لغرض أن تكون المجتمعات أكثر أماناً، وأمناً، وتنافسية، وإنتاجية، وكفاءة، وشفافية ومساءلة، على الرغــم من أنه يتم القيــام بذلك من خلال العمليات التي تراقــب وتفرض النظام، وتكبح، وتقنع، وتجبر، وتســتغل الناس. فهناك توازن دقيق، في هذه الحالة، بين استخدام البيانات بطرق تحريرية وتمكينية، وبين اســتخدام البيانات من أجل تحقيق أغراض الفرد الشخصية وإلحــاق الــضرر بالآخرين، أو اســتخدام البيانات بطرق تتعارض مع رغبــات هؤلاء الذين تمثلهم هذه البيانات. ولا يتعلق الأمر، مع ذلك، بأن تســتخدم البيانات ببســاطة في أشياء جيدة أو أشــياء سيئة، فالمشــكلة أعقد من ذلك بكثير. وفي غالب الأحيان فإن النتائج التي تظهر على أنها متعارضة يتم تجميعها سوياً لدرجة أن الأشخاص يمكن تحريرهم أو إجبارهم في الوقــت نفســه - وهم يحصلون على منافع شــخصية في الوقت ذاتــه الذي يقعون فيه في مصيدة النظام الذي يســعى إلى تحقيق مكاســب من وراء مشاركتهم. وفي مصطلحات Althusser (1971)، تعمل مثل هذه الترتيبات من خلال الاســتجواب، وتوريط الناس عن طريق منطقها وذلك من خلال الإقناع والحوافز. فعلى سبيل المثال، تمنح بطاقات الولاء في الأســواق للعملاء خصومات في الوقت ذاته الذي تعمل هذه البطاقات على تأســيس الولاء تجاه المتجر وتكوين قدر من البيانات التي يتم استخدامها من أجل محاولة بيع المزيد من البضائــع لهؤلاء العملاء، مما يؤدي إلى زيادة الأرباح. وبالمثل، فإن ثمن حصولك على الأمان من الهجمات الإرهابية هو المراقبة التي تنتهك خصوصيات جميع الأفراد في المجتمع، حيث يحصل المواطنون على الأمان مقابل خصوصيتهم.



إن تكويـن البيانـات والعمل الذي تقوم به مثل هذه البيانات محاط بشـكل جوهري بمخـاوف أخلاقيـة، واجتماعية، وسياسـية. ومثل هذه المخاوف كانـت لوقت طويل مثار إدراك ومجادلات داخل المنتديات العامة والمنتديات العلمية، مما أدى إلى صياغة مجموعة كبيرة من الإرشـادات الأخلاقية والمهنيـة والتشريعات التي تحدد كيف يتـم إنتاج، إدارة، مشاركة، واستخدام البيانات. فليس هناك أي شكل من أشكال إنتاج وتحليل البيانات خالٍ من مثل هذه المخاوف، وحتى المشـاريع العلمية التي يبدو من الناحية الظاهرية أنها غير خطرة ومحايدة، يمكن أن تثير أسـئلة أخلاقية ويكون لها توابع اجتماعية وسياسـية. فعلى سـبيل المثال، ربما يبدو ظاهريًا أن قياس ونمذجة الأنظمـة البيئية يكون مقتصراً على فهم الظواهر الطبيعية، ولكن من الممكن أن يتم استخدام ذلك أيضًا في تشكيل السياسات التي تؤثـر في حياة الناس داخل مثل هذه الأنظمة. إن مثل هذه الحوارات الأخلاقية مسـتمرة وتتطور واحدة تلو الأخرى مع المجال الاسـتطرادي الأشمل الذي يتعلق بالاتجاهات والآراء الاجتماعية، وتطور التقنيات الجديدة. وفيما يتعلق بالأخيرة، فإن توسيع البيانات الصغيرة داخـل البنى التحتية للبيانات، وإنشـاء سـوق البيانات الهائل، وفتح البيانات المؤسسـية، وتدفقات البيانات الكبيرة يطرحان أسـئلة أساسـية أخلاقية، واجتماعية، وسياسية، إذ إنها تعدل بشكل جذري مشهد البيانات. إن حجم وتنوع البيانات التي يتم تكوينها عن جميع مجالات الحياة اليومية والعالم الذي نعيش فيه اليوم ينموان بشـكل مطرد، وهذه البيانات هي أكثر من أي وقت مضى من حيث مناسـبة توقيتها، وتصميمها، وشـموليتها، وترابطها. فـما كان في الماضي خاصـاً أو مجهولاً أصبح الآن وبشـكل متزايد مكشـوفاً أمام مجموعة متنوعـة من المصالح، وعمليات اتخـاذ القرار داخل الحكومـات، وأصبحت الأعمال توجه بالبيانات بشكل أكبر، ومدعومة بالدليل وأكثر اعتمادًا على التقنية (تكنوقراطية).

يناقش هذا الفصل مجموعة مختارة من المخاوف الأخلاقية، والاجتماعية، والسياسـية، والقانونيـة التـي تثيرها ثـورة البيانات، وهذه المخاوف تشـمل مراقبـة البيانات وتعقب البيانـات، والخصوصية، وأمـن البيانات، والتشـخيص، والتصنيـف الاجتماعي والخطوط الحمراء (Redlining) والتي تعني ممارسـة الحرمان أو زيادة التكلفة في الخدمات لمناطق لهـا غالبية عرقية محددة بناء عـلى البيانات، والتحكم في الانسـلال، والحوكمة التوقعية، وحوكمة المؤسسات، والاعتماد التام على التقنية وحوكمة الشركات والإغلاق التقني، والملكية



وحقوق الملكية الفكرية. فالفصل يناقش كيفية التفكير في كل قضية من هذه القضايا، مع الآراء المتباينــة داخل كل علم على حدة وبين العلوم المختلفة بعضها البعض، وكذلك داخل وبين الشركات والحكومات والمجتمع المدني، ومن لديه أجندات مختلفة، ومصالح مكتسبة، وحساســيات سياســية. ولذا، فإنه لا توجد إجابات ســهلة لحل هذه القضايا التي يناقشها هذا الفصل، في حين أن الحلول تتكون دوماً من تســويات وحلول وسط. ومن المؤكد، على مــا ذكرنا، أنه كلما تنتشر ثــورة البيانات، فمن المرجح أن تزداد حدة المناقشــات والجدال بشــأن هذه المخــاوف، وخصوصاً عند القيام بمحاولات من أجــل صياغة تشريعات جديدة للتعامل مع التطورات التقنية التي تساعد على انبثاق طرق جديدة لإنشاء وتعزيز وتحليل البيانــات، ومن ثم تــؤدي إلى ظهور قضايا جديــدة مما يجعل من التشريعــات والقوانين القديمة تشريعات وقوانين عفا عليها الزمن.

تعقب البيانات ومراقبة البيانات (Data Shadows and Dataveillance):

كما ناقشــنا في الفصول السابقة، يتم تكوين كمية هائلة من البيانات في الوقت الراهن وذلــك فيما يتعلق بالمواطنين في جميع ولايات الدولة. إن إنتاج وتحويل مثل هذه الكمية من البيانات ينمو ويتزايد باطراد، حيث تسعى الدولة والشركات ومنظمات المجتمع المدني بشــكل فعال للحصول على البيانات المتعلقة بمواطنيهم وعملائهم والأعضاء المشــتركين في أنشــطتهم. وفي الواقع، أصبح من الصعوبة بمكان أن يــؤدي الفرد جزءًا من حياته اليومية دون أن يــترك أثرًا لذلك بســبب الدور غير المبــاشر للتقنيات الرقمية والبرامج واســتخدام المعرفات الدالية (Kitchin and Dodge 2011). فحتى ولو لم يقم المشــتري باســتخدام بطاقتــه الائتمانية لشراء البضائع من المتاجر، إلا أن حضور هذا المشــتري قد تم تســجيله من خلال كاميرات المراقبة، وحتى لو اســتخدم الشخص اسم مستخدم مجهول على وسائل التواصــل الاجتماعي فإنه يتم تســجيل عنوان بروتوكول الإنترنــت (IP Address) وعنوان التحكم بالوصول للوســائط (MAC Address). ومن ثم فنحن نترك آثار البيانات بشــكل روتيني في أعقابنا، على الرغم من أننا في الغالب لا نمتلك السيطرة الكاملة على شكل ومدى وكيفية استخدام هذه الآثار.



وكما أشـار (2011) Koops، على سـبيل المثال، فإن هيئة حماية البيانات الهولندية قـدرت أن المواطن الهولندي العادي منـدرج في نحو ٢٥٠ – ٥٠٠ قاعدة بيانات، في حين أن المواطنين الأكثر نشاطاً من الناحية الاجتماعية مندرجون في نحو ١٠٠٠ قاعدة بيانات. ولا تشـتمل قواعد البيانات هذه على الآثار الرقمية للمواطنين فحسـب (البيانات التي يتركها المواطنون أنفسهم في أعقابهم)، ولكنها تتضمن أيضًا بيانات الصور التخيلية (Data Shadows) الخاصـة بالأفراد (البيانات عن المواطنين التي ينشـئها الآخرون)، وهي توفر بشـكل متزايد عملية تعقب البيانات الخاصة بالموقـع والتفاعلات والمعاملات عبر المكان والزمـان (Clarke 1994a; Lyon 2007; Dodge and Kitchin 2005; Koops 2011). وعـلاوة على ذلك، ففي حين أن الآثار الرقمية والصور التخيلية للبيانات سريعة الزوال ولا تدوم إلا فترة محدودة للغاية، فإن النسخة الرقمية منهما تدوم بشكل كبير للغاية ويمكن تخزينهـا في قواعد البيانات لفـترة غير محدودة (Koops 2011)، مما يؤدي إلى إنشـاء عـالم من الممكن له أن يتذكر جميع نقاط البيانات ولا ينسى على الإطلاق (Dodge and Kitchin 2007b). فهذه الآثار الرقمية والصور التخيلية للبيانات يتم تفتيتها وتقسيمها ونشرها عبر الكثير من المنظمات والخوادم، وهي تكون عرضة للدمج والتقسـيم (Raley 2013). وفي أفضل الأحوال، تشـكل الآثار الرقمية والصور التخيلية للبيانات القليل من الـرؤى – رؤى محدودة من زوايا مرتفعة جزئية من مواقف محددة مع إسـقاط الرؤى المحـددة (Amin and Thrift 2002) – وذلك بدلاً مـن تمحيص هذه الآثار والتخيلات سـوياً من أجل تكوين رؤية شـاملة وعامة تحيط بكل شيء. وعلى الرغم من ذلك، وإذا وضعنا في اعتبارنا ارتباطية البيانات وقيمة تضخيم البيانات، فإن الاتجاه السـائد سيكون تجاه تجميع وجهات النظر المحدودة من أجل تكون تصور أكثر قوة.

وإجمالاً، توفر الآثار الرقمية والصور التخيلية للبيانات سجلاً مفصلاً بحياة المرء اليومية، وأنماط اسـتهلاكه، وعمله، وسـفره، واتصالاته، ولعبـه، وتفاعلاته مع المنظـمات، وأفكاره واهتماماته. وليس الأفراد فحسب هم من يقعون تحت طائل هذه المراقبة الشاملة، ولكن هذه المراقبة تشمل أيضًا الأشياء والمؤسسات والتعاملات والمناطق (Dodge and Kitchin 2005). ولم يحدث أبداً في السـابق أن يتم تكوين مثل هذه الكمية من البيانات عن حياة الناس الاقتصادية والاجتماعية بسهولة بالغة، وأن يتم إتاحة هذه البيانات على نطاق واسع



للغاية من خلال أسـواق البيانـات والبنى التحتية للبيانات ومبـادرات البيانات المفتوحة. وفي الواقـع، فإن اكتسـاب الوصول إلى البيانات الحساسـة ليس مقتـصراً على الحكومات والشركات والمنظمات، وذلك بسـبب أن مراقبة النظراء أصبحت أمراً أكثر انتشـاراً بشـكل نسـبي دون معرفة أو دون الحصول على إذن الشـخص الخاضـع للمراقبة (Andrejevic 2007). وبالطبع، كان هذا الأمر متاحاً على الدوام من خلال التطفل والقيل والقال، ولكنه أصبح أكثر سـهولة في الوقت الحالي من خلال الإنترنت سواءٌ كان هذا من خلال البحث عن البيانات أو شراء تلك البيانات، أو من خلال اسـتعراض سـجلات الأجهزة الرقمية كالهواتف الخلوية، وبرامج تصفح الإنترنت، أو برامج الملاحة بالأقمار الصناعية في السـيارات... وغير ذلك، أو من خلال وضع أجهزة مراقبة رخيصة نسبياً مثل الكاميرات الدقيقة المصغرة. وكما ناقشـنا في الفصل الثاني، فإن سـماسرة البيانات يسمحون بدراسـة تواريخ الأشخاص وذلك فيما يتعلق ببعض المسـائل مثل الحالة الاجتماعية، والإفـلاس، وملكية العقارات، والتاريخ الوظيفي، والسجلات الجنائية، والحالة الضريبية، ودعاوى العنف الجنسي أو العنف الأسري، والخلفية التعليمية... وغير ذلك.

ومن ثم فإن طوفان البيانات مكشـوف أمام مراقبة البيانات والاستغلال، فمُراقبة البيانات، تُعد أحد أساليب المراقبة التي يتم تفعيلها من خلال فرز وتدقيق مجموعات البيانـات وذلـك بغرض التحديـد والمراقبة، والتتبـع، والتنظيم، والتنبـؤ، والتوجيه (Clarke 1988; Raley 2013)، وتعمل مراقبة البيانات على تحويل الكميات الهائلة مـن الأفعال والأفـكار التي تمثلها هذه البيانات إلى صـورة واضحة ومقروءة ونماذج واضحـة للحكم على الأشـياء (Curryet al. 2004: 359)، وهي مكون أسـاسي من الأشكال الحديثة للحوكمة والحاكمية. وفيما يتعلق باستخدام البيانات، وكما ناقشنا في الفصل السـابع، يتم اسـتخدام البيانات من أجل توصيف واسـتهداف الأشخاص وذلـك بهدف زيادة الأرباح. ومـن ثم ففي غالب الأحيان يتم إعادة تحديد الغرض من البيانات واسـتخدامها في أغراض لم تكن مقصودة على الإطلاق عند تكوين هذه البيانـات، ومن ثم فإن مراقبة الآثار الرقمية والصور التخيلية للبيانات يثيران العديد مـن المخاوف الإضافية وخاصة فيما يتعلق بقضايا مثل الخصوصية، والتنميط والفرز الاجتماعي، وإقرار أساليب متعددة للحوكمة.

الخصوصية (Privacy):

الخصوصيـــة هي حالة يتوقعها ويقدرهـــا الكثير من الناس، وهي تُعَدُّ حقاً من حقوق الإنســـان الأساســـية، وهي من الحقوق المقدســـة في الكثير من القوانين الوطنية والدولية. وفي الولايـــات المتحـــدة، فإن الخصوصية مغطاة بشـــكل كبير في القوانين المســـماة قوانين الخصوصيـــة، أما في الاتحـــاد الأوروبي، فهي تقـــع ضمن نطاقات قوانـــين حماية البيانات (Minelli et al. 2013). والخصوصيـــة مصطلح متعدد الأبعاد فيما يتعلق بمعناه، وهو يســـتخدم في غالب الأحيان في طرق معتمدة على الســـياق، ولكـــن على وجه العموم هو يشـــير إلى الممارسات المقبولة فيما يتعلق بالوصول إلى والإفصاح عن المعلومات الشخصية والحساســـة (Elwood and Leszczynski 2011). ومن الوســـائل المستخدمة في إيضاح الأبعـــاد المتعددة للخصوصية دراســـة الوســـائل العديـــدة التي يمكن مـــن خلالها انتهاك الخصوصية والأضرار المرتبطة بهذه الانتهاكات، (Solove 2006، انظر الجدول رقم 10.1). فالـــذي يتضح لنا من خـــلال هذا التصنيف هو أن انتهـــاك الخصوصية يمكن أن يكون له العديـــد من التأثيرات على الرفاهية الجســـدية والعاطفية للأفـــراد، ويجعل هؤلاء الأفراد عرضةً للأنشـــطة المضرة التي يقـــوم بها الآخرون، ويجعلهم كذلـــك عرضة لاختلال توازن السلطات غير المتماثل.

ويوجد بعض الشكوك بأن مفهوم الخصوصية يتغير. وكما تمت الإشارة أعلاه، فإن الناس عرضة الآن لمستويات من الفحص والتدقيق وكذلك أساليب المراقبة أكثر من أي وقت مضى. فالمهـــام التي كانـــت تعد في الماضي خاصة إلى حد كبير، يتم الآن مراقبتها وتســـجيلها. على سبيل المثال، فإن البرامج التي كان يشاهدها الشخص على التلفاز لم تكن معروفة في الماضي إلا لمـــن هم حاضرون مع الشـــخص ذاته في الغرفة ذاتها، في حـــين أنه يمكن لشركات التلفاز الآن مراقبة ما الذي يتم مشـــاهدته على كل تلفاز على حدة وذلك بســـبب اســـتبدال البث التلفزيوني وحيد الاتجاه بالتواصل ثنائي الاتجاه. وما يشـــتريه المرء كان في الســـابق معروفاً للعميل والشخص الذي يخدم هذا العميل فحسب والآن أضحت هذه المعلومات التي يتم جمعها بشـــكل روتيني من خلال إدراج النقود الرقمية في عملية الدفع متاحة لشرائح أكبر. والأماكن التي كان يســـافر إليها الفرد لم تكن سابقاً معروفة إلى حد كبير لأي شخص ما عدا

المســافرين ومرافقيهم، ولكن في الوقت الراهن أضحى العديد من الأنظمة تكشــف موقع الشــخص والطرق التي اجتازها كنظم الملاحة باســتخدام الأقــمار الصناعية (GPS) المثبتة في الســيارات أو أجهزة الهواتف الخلوية، ونظام تثليث الإشارات في أبراج شبكات الهاتف، وأنظمة النقل العام التي تعتمد على اســتخدام شريحة تحديد الهوية باســتخدام موجات الراديو (RFID) في أنظمة أو أكشــاك دفع رســوم المرور. ومثل هذه البيانات التي تنتجها هذه الأنظمة يتم إعادة تجميعها ونشرها وبيعها.

الجدول رقم (١ - ١٠)

تصنيف الخصوصية

النطاق	انتهاك الخصوصية	الوصف
جمع المعلومات Information Collection	المراقبة	مشــاهدة أو الاستماع إلى، أو تســجيل الأنشطة التي يقوم بها الشخص
	الاستجواب	الأشــكال المتعددة من الاستجواب أو التقصي من أجل الحصول على المعلومات
معالجة المعلومات Information Processing	التجميع	تجميع أجزاء مختلفة من البيانات عن شخص ما
	التعريف	ربط المعلومات بأشخاص معينين
	عدم الأمان	الإهــمال في حماية المعلومات المخزنة من التسريبات والوصول إليها بطريقة غير ملائمة
	الاستخدام الثانوي	يتم استخدام المعلومات التي تم تجميعها لغرض معين في أغراض أخرى دون الحصول على موافقة الشــخص موضوع البيانات
	الاستبعاد	الإخفــاق في الســماح للشــخص موضــوع البيانــات بمعرفة البيانات التي يمتلكها الآخرون عنه والمشــاركة في معالجــة واســتخدام هذه البيانات، وهذا يشــمل الإقصــاء والمنع من التمتع بالمقــدرة على الوصول إلى هذه البيانات وتصحيح الأخطاء الواردة فيها.

نشر المعلومات Information Dissemination	انتهاك السرية	إخلاف الوعد في المحافظة على سرية بيانات الشخص
	الإفصاح	الكشف عن المعلومات عن شخص ما والتي تؤثر في الطرق التي يحكم بها الآخرون على شخصيته
	الكشف	الكشف عن عري أو حزن أو الوظائف الجسدية لشخص ما
	زيادة إمكانية الوصول	زيادة إمكانية الوصول إلى المعلومات
	الابتزاز	التهديد بالكشف عن معلومات شخص ما
	الاستيلاء	استخدام هوية الشخص موضوع البيانات من أجل خدمة أهداف أو مصالح شخص آخر
	التشويه	نشر معلومات خاطئة أو مضللة عن شخص ما
الانتهاك Invasion	التطفل	الأعمال الانتهاكية التي تزعج هدوء أو انعزال شخص ما
	التدخل المتعلق بالقرارات	التدخل في قرارات الشخص موضوع البيانات التي تتعلق بشئونه الخاصة.

المصدر: تم تجميعه من Solove (2006).

وعلاوةً على ذلك، فنحن نعيش في عالم أكثر انفتاحاً وشفافية بكثير عما اعتدنا عليه سابقاً. فالمعلومات التي كانت تُعدُّ خاصةً في السابق يتم مشاركتها الآن بشكل أكثر حرية كالسير الذاتية من خلال موقع لينكد إن (LinkedIn)، والصور والفيديوهات العائلية من خلال مواقع فليكر (Flickr) وإنستجرام (Instagram) ويوتيوب (YouTube)، والقصص الشخصية والعائلية من خلال موقع فيسبوك (Facebook) والمدونات، والأفكار والمعتقدات الشخصية من خلال تويتر (Twitter) وغرف الدردشة والتعليقات على الإنترنت. فما كان يتم مشاركته في السابق مع عدد لا يتجاوز أصابع اليد الواحدة من الأشخاص كأفراد العائلة، والأصدقاء المقربين، وأرباب العمل في المنزل أو المقهى المحلي أو مكتب للموارد البشرية يتم نشره الآن على المستوى العالمي مما يسمح لأي شخص بمشاهدته وتعديله (Minelli et al. 2013).

وعلى الرغم مما ذُكر ســابقاً، لا يتصرف كل شــخص بســهولة وبحرية فيما يتعلق ببياناته الشخصية، حتى ولو كان الشخص يتصرف على هذا النحو، فإنه لا يتوقع بالضرورة أن هذه البيانات ســوف يتــم انتقاؤها، ومعالجتها، وتجميعها، وبيعهــا. وهم لا يتوقعون أيضًا أن تطبيقات البرامج التي لا يوجد بها أي عنصر من عناصر وسائل التواصل الاجتماعي، والتي يعتبرونها تطبيقات خاصة ربما تقوم بتكوين ونقل البيانات إلى مطوري تلك التطبيقات، أو أن بعض الشركات ربما تحاول أن تقوم بالهندسة العكسية لإستراتيجيات حماية الخصوصية مثــل إخفاء الهوية من أجل إنشــاء ملفات شــخصية موحدة، ومع ذلك، فــإن هذا الأمر يحدث بانتظام. على ســبيل المثال، وجدت صحيفة وول ســتريت جورنال (Wall Street Journal) في اختبــار أُجري على 101 تطبيق للهواتف الذكية، أن 56 تطبيقاً نقلت معرف الجهــاز المتفرد للهاتــف إلى شركات أخرى وذلك دون معرفة المســتخدم أو الحصول على موافقته، وأن 47 تطبيقاً أرسلت موقع الهاتف، وأن 5 تطبيقات أرسلت البيانات الشخصية للمســتخدمين، وأن 45 تطبيقاً لم يكن بها أي ارتباط مع سياســات حماية الخصوصية التي يمكن للمستخدمين مشاهدتها (Efrati et al. 2011). كما وجدت شركة الثقة الإلكترونية (TRUSTe) أن ١٩٪ فقط من أعلى ٣٣٤٠ تطبيقاً مرتبطون بسياسة الخصوصية وأن متاجر شركتي أبل وجوجل يشــترطان ضرورة اشــتمال التطبيقات على هذه السياســة (Coterill 2011). ومــع ذلك، وكما أشــار Gralla et al (2011)، يمكــن للتطبيقات أن تتتبع وتنقل "عاداتــك على الشــبكة، وأن تبحث في قائمة جهــات الاتصال لديــك، وأن تجري مكالمات هاتفية دون أن تعرف، وأن تتتبع موقعك، وأن تفحص ملفاتك، وغير ذلك الكثير".

إن مشهد الخصوصية الآن في حالة تغير مستمر، وهو يخيب كلاً من التوقعات القانونية والاجتماعية، وبالنســبة للبعض، فإن مفهوم الخصوصية قد تعطل بشكل كبير (Rambam 2008; Rubenking 2013). ويرى البعض أنه من الصعوبة بمكان المحافظة على الخصوصية في التطبيق العملي، كونها تحجب وتخفي تجربة المستخدم، وهي معوق اقتصادي، ولا يبدو أن معظم الناس يعارضون أن يتم التنقيب في بياناتهم، وإذا لم يكن لديك شيء لتخفيه، فما المشكلة في أن يتم معرفة بياناتك؟ (Raley 2013: 126; Solove 2007). وبالنسبة للبعض الآخر، فإن الخصوصية حق يتعين حمايته لأنه من الحقوق التأسيســية لـ "المواطنة المطلعة والمتأملة" وهو مكون أساسي أيضًا لحرية التعبير (Cavoukian 2009). وعلى هذا، يُرى أن



الخصوصية "سـمة ضرورية وتأسيسية للأنظمة السياسية الديموقراطية المتحررة" (Cohen 2012: 2). فـإذا تم إنكار الأشـكال المتعددة من الخصوصية التـي تم إيجازها في الجدول رقم (١-١٠)، فإن الناس سـيكونون عرضة لأنوع مختلفة من الأضرار الشـخصية وأشـكال الحوكمة الأكثر ضرراً. ويبدو أن جميع المعلقين يؤيدون وجهة النظر القائلة بأن تشريعات الخصوصية لم تعد ملائمة للغرض التي شُرعت من أجله، وأنه يتعين تحديث تلك التشريعات لكي تلائم الأوقات التي نعيشها الآن.

الجدول رقم (٢ - ١٠)

مبادئ الممارسة العادلة للمعلومات

المبدأ	الوصف
الإخطار (Notice)	يتم إخطار الأفراد بأن بياناتهم يتم إنشـاؤها ويتم إعلامهم بالغرض الذي سيتم استخدام هذه البيانات فيه
الاختيار (Choice)	يتمتع الأفراد بحرية اختيار المشاركة أو يؤثرون عدم المشاركة وذلك فيما يتعلـق بما إذا كانوا يرغبون في اسـتخدام بياناتهم أو الإفصاح عنها وكيفية القيام بذلك
الموافقة (Consent)	يتـم إنشـاء البيانات والإفصاح عنها فحسـب بعـد موافقة الأفراد المعنيين
الأمن (Security)	يتم حماية البيانات من الفقد، وإسـاءة الاسـتخدام، والوصول غير المصرح به، والإفصاح عنها، وتعديلها أو تدميرها.
التكامل (Integrity)	تكون البيانات موثوقة ودقيقة وكاملة وحالية
الوصول (Access)	يمكن للأفراد الوصول إلى بياناتهم الشخصية والتحقق منها والتثبت منها
المساءلة (Accountability)	يكـون مقتني البيانات مسـئولاً عن ضمان تحقق المبادئ السـابق ذكرها وأن يكون لديه الوسائل اللازمة من أجل ضمان الالتزام بهذه المبادئ

المصدر: (Minelli et al. (2013: 156.



الجدول رقم (٣ - ١٠)

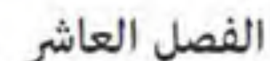
أنواع المعلومات المحمية

المعلومات الشخصية المعرفة (PII): هي أي معلومات تُعرف شخصاً ما بصورة مباشرة أو غير مباشرة	المعلومات الحساسة: هي أي معلومات يُمكن للكشف غير المصرح لها أن يسبب الإحراج أو الضرر لشخص ما	المعلومات الأخرى التي يمكن استخدامها لتخمين هوية شخص ما
الاسم	العرق / الانتماء العرقي	الأشياء المفضلة
العنوان البريدي / الرمز البريدي	الآراء السياسية	هوية ملف تعريف الارتباط (cookie)
البريد الإلكتروني	المعتقدات الدينية / الفلسفية	عنوان بروتوكول الإنترنت الثابت
رقم الجوال / الهاتف	عضوية النقابات التجارية	
رقم الضمان الاجتماعي	المعلومات الصحية / الطبية	
رقم رخصة القيادة	الحالة الاجتماعية / الحياة الجنسية	
رقم الحساب المالي (البنكي)	العمر	
رقم بطاقة الصراف / بطاقة الائتمان	الجنس	
	السجل الجنائي	

المصدر: مقتبس من Minelli et al. (2013: 159).

وفي الوقت الراهن، فإن تشريعات الخصوصية مبنية حول الحقوق الشخصية والموافقة المتعلقة بإنشاء، واستخدام، والإفصاح عن البيانات الشخصية (Solove 2013). وهي تتكون من سبعة مبادئ عالمية للخصوصية (Minelli et al. (2013: 156، (انظر الجدول رقم ٢ - ١٠) وفي إطار هذه المبادئ، فإن الأفراد، من الناحية النظرية، يتم منحهم الحق للتحكم في بياناتهم الشخصية ومنح موافقتهم للآخرين فيما يتعلق بهذه البيانات. وعلاوة على ذلك، فإنه يتعين على مقتنيي البيانات الالتزام بقوانين حماية البيانات التي تقيد الإفصاح عن المعلومات الشخصية





المعرفـة (Personally Identifiable Information- PII)، والمعلومـات الحساسـة، والبيانات الأخرى التي يمكن استخدامها للاستدلال على الشخصية (انظر الجدول رقم ٣-١٠)، وتؤيد تقليص البيانات بحيث يقتصر إنشـاء البيانات على البيانات الضرورية لتحقيق غرض معين (Tene and Polonetsky 2012). وعـلى وجه العموم، فإن الحل لاشـتراطات المعلومات الشـخصية المعرفة (PII) هو استخدام أسـاليب إخفاء الهوية مثل عدم تحديد الهوية (إخفاء المعلومات الشخصية المعرفة)، والأسماء المسـتعارة، والتجميع، بالإضافة إلى التشفير، والتخزين الآمن، والقيود الخاصة بالوصول إلى المعلومات (Coterill 2011).

وأخـيراً، تجـدر الإشـارة إلى وجود العديـد من المشـاكل الإدراكية والبنيويـة المرتبطة بتشريعات الخصوصية الحالية.

وتحـاول الدول بصورة روتينية التحايل على قوانـين الخصوصية وذلك من أجل أغراض الأمن وجمع المعلومات الاسـتخباراتية، بحسب ما تم الكشـف عنه مؤخراً بواسطة برامج تحليل وجمع البيانات السرية في الولايات المتحدة الأمريكية، والمملكة المتحدة، والعديد من الحكومات الأخرى (انظر الفصل السابع). كما أن الكثير من الشركات لا تشعر بأنها مجبرة على تقديم سياسة الخصوصية للأفراد وذلك لأن هذه الشركات تدعي أنها تجمع معلومات مجهولـة المصـدر، ومن ثم فهي لا تقع تحت طائلة ممارسـات المعلومات العادلة. أو ربما تقدم هذه الشركات سياسة الخصوصية, وذلك عند الاستهلال الأولي للالتزامات (على سبيل المثال، عند تثبيت برنامج ما أو الاشـتراك في خدمة ما) وتكون هذه السياسـة مصاغة بلغة معقدة وغامضة، وعادةً ما تشـتمل على الاحتفاظ بالحق في تعديل هذه السياسة في وقت لاحق دون تشاور إضافي مع الشخص (Rubinstein 2013).

وفيـما يتعلق بالبيانـات التي تم تكوينها بدون معرفة الشـخص، فمن الصعوبة بمكان اكتشـاف مثل هذه الممارسـات ومواجهتها من الناحية القانونية. أما في حالة الخصوصية ذات الإدارة الذاتيـة، وكـما شرح Solove (2013)، فهـي ببسـاطة مرهقـة للغاية وبالغة الصعوبـة على الأفراد كونها تتطلب قيامهم بتنظيـم خصوصيتهم عبر الكثير من الكيانات، والموازنـة بين التكاليف والمنافع المرتبطة بالموافقـة على الشروط والبنود دون معرفة كيف سيتم استخدام هذه البيانات في الوقت الحالي وفي المستقبل، وأيضًا تقييم التأثيرات الشاملة



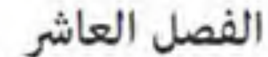
والتراكمية لبياناتهم التي سيتم دمجها مع مجموعات البيانات الأخرى. وبناء على ذلك، أشار Solove (2013) إلى أن:

(1) الناس لا يقرؤون سياسات الخصوصية؛ (2) إذا قرأ الناس سياسة الخصوصية، فإنهم لا يفهمونها؛ (3) إذا قرأ الناس سياسة الخصوصية وفهموها، فإنهم في الغالب يفتقدون الخلفية الكافية لاتخاذ اختيار مبني على المعرفة؛ (4) وإذا قرأ الناس سياسة الخصوصية وفهموها، وكان بإمكانهم اتخاذ اختيار واعٍ ومبني على المعرفة، فإن اختيارهم ممكن أن ينحرف وذلك من جراء العديد من الصعوبات المرتبطة باتخاذ القرار (Solove 2013: 1888).

ومن ثم فإن الموافقة تكون في غالب الأحيان ممارسة عديمة الجدوى (Rubenstein 2013: 2)، وذلك عندما يقوم الأفراد عن غير قصد بالتخلي عن الحقوق دون إدراك لمدى أو عواقب الفعل الذي يقومون به. ومن ثم، فإن سياسات الخصوصية في غالب الأحيان تخدم إخلاء المسئولية للشركات بصورة أكبر مما تخدم عملية ضمان الخصوصية للمستهلكين، (Tens and Polonestky 2012). وعلاوة على ذلك، فمن الواضح أنه في عصر البيانات الكبيرة، فإن إستراتيجية إزالة التعرف على الأشخاص توفر شكلاً ضعيفاً فحسب من أشكال الخصوصية لأن من الممكن القيام بالهندسة العكسية للعملية وذلك من خلال التنقيب في مجموعات البيانات وتجميعها (Narayanan and Shmatikov 2010). فعلى سبيل المثال، يشرح نيف سينجر (Niv Singer) من شركة تراكسك (Traxc) (والمقتبس في Minelli et al. 2013: 31) عملية إزالة التعرف على الأشخاص كما يلي:

> يمكن في بعض الأحيان أن يكون من الصعوبة بمكان توحيد الملفات الشخصية الاجتماعية لمستخدم واحد والذي ربما يستخدم العديد من الأسماء أو تعاملات مختلفة في كل واحدة من شبكات التواصل الاجتماعي، ولذا فلقد قمنا ببناء خوارزميات تقوم بالتنقيب في العوامل الأساسية التي تشتمل على محتوى المشاركات والموقع على سبيل المثال لا الحصر، وذلك من أجل توفير توحيد للهوية بشكل فعال إلى درجة كبيرة.





ونتيجة لهذه المخاوف، يدرس الاتحاد الأوروبي منذ فترة سياسات حماية البيانات (انظر المفوضية الأوروبية ;European Commission 2012، Hon et al. 2011; Koops 2011 Rubinstein 2013). وتشتمل اقتراحات الاتحاد الأوروبي على أن تكون الموافقة صريحة وليست ضمنية أو مفترضة، وأن يكون وصول الأفراد إلى البيانات الخاصة بهم سهلاً وشاملاً، وأن يكون للأفراد الحق في قابلية النقل للبيانات بمعنى المقدرة على نقل البيانات الشخصية من مزود خدمة إلى مزود آخر، والحق في أن يتم نسيان الشخص، حيث يمكن للأفراد أن يطلبوا حذف بياناتهم إذا لم تكن هناك أي أسس مشروعة للاحتفاظ بهذه البيانات، وأن تنطبق هذه القواعد أيضًا على الشركات خارج الاتحاد الأوروبي إذا كانت هذه الشركات عاملة في سوق الاتحاد الأوروبي وتقدم خدماتها لمواطني دول الاتحاد الأوروبي (European 2012 Commission). وبالمثل، تقدمت مفوضية التجارة الفيدرالية في الولايات المتحدة الأمريكية باقتراح يتضمن ثلاثة تعديلات من أجل توفير حماية أكثر فعالية للخصوصية: الخصوصية من خلال التصميم حيث يتم تأسيس الخصوصية بشكل متأصل في كل مرحلة من مراحل تطوير المنتج؛ منح الشركات والمستهلكين خيار مبسط يمنحهم المقدرة على اتخاذ القرارات بشأن بياناتهم الخاصة، وهذا يشمل تطبيق آلية لعدم التتبع والحصول على الموافقة الصريحة فيما يتعلق بالبيانات الحساسة، أو الحصول على موافقة قبل استخدام البيانات في غرض مختلف بشكل مادي عن الغرض الذي تم من أجله تكوين هذه البيانات؛ وشفافية أكثر فيما يتعلق بالبيانات وجمعها واستخدامها، وهذا يشمل الوصول المعقول إلى البيانات من قبل هؤلاء الذين تمثلهم تلك البيانات، مع منحهم المقدرة على تصحيح أو حذف البيانات (مفوضية التجارة الفيدرالية Federal Trade Commission 2012). وتنشد إستراتيجية الخصوصية من خلال التصميم تطبيق إستراتيجية مختلفة عن الالتزام التشريعي والتنظيمي، وهي تقترح أن تكون الخصوصية هي الوضع الافتراضي للعملية؛ أي نظام يسعى بفعالية وبوضوح إلى ضمان الخصوصية (Cavoukian 2009). ومن شأن المبادئ السبعة المذكورة في الجدول رقم (٤ - ١٠) أن تعزز هذا المنهج.



الجدول رقم (٤ - ١٠)

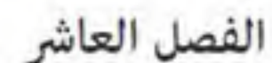
المبادئ السبعة الأساسية للخصوصية من خلال التصميم

المبدأ	الوصف
استباقية وليست تفاعلية؛ الوقاية وليس العلاج	ينبغي أن تسعى نظم تكنولوجيا المعلومات لتوقع الخصوصية تحديد المخاوف المتعلقة بالخصوصية وتجنبها بدلاً من السعي إلى حل المشاكل الناتجة عن مخالفات الخصوصية بعد حدوثها
الخصوصية هي الوضع الافتراضي	الخصوصية محمية تلقائيا ولا تتطلب أي عمل من جانب الأفراد
الخصوصية مدمجة في التصميم	حماية الخصوصية هي سمة أساسية للتصميم والهندسة الهيكلية لأنظمة تقنية المعلومات وليست ميزة إضافية متدنية الأهمية
الفعالية الكاملة - عائد إيجابي وليس محصلة صفرية	جميع المصالح والأهداف المشروعة يتم استيعابها، بدلا من أن يكون هناك مقايضات بين الخصوصية واعتبارات أخرى مثل الأمن
الأمن من النهاية للنهاية-حماية كاملة طوال دورة حياة النظام	تكون الخصوصية جزءًا لا يتجزأ من النظام من بداية إطلاقه إلى حين التخلص منه
الوضوح والشفافية - فلتكن مفتوحة على الدوام	تكون جميع الأجزاء المكونة للنظام وعملياته واضحة وشفافة للمستخدمين والشركات الموفرة على السواء وتخضع لتدقيق مستقل
احترام خصوصية المستخدم - أي جعلها تتمحور حول المستخدم	يجب أن يبنى النظام حول الأفراد، وأن يحمي مصالحهم، وأن يكون مُمكنًا لهم

المصدر: (Cavoukian 2009).

لقد اقترح العديد من العلماء المتخصصين في الخصوصية وأيضًا العديد من مجموعات الضغط الصناعية مناهج متعددة لكيفية التعامل مع قضية الخصوصية، فالعلماء المختصون بموضوع الخصوصية سبق لهم التطرق للمواضع التالية: الإصلاح التشريعي المصمم من أجل حماية حقوق المواطنين، والذي يشمل بعض القضايا كالحق في الصفحة البيضاء، وتواريخ الانتهاء للبيانات (Koops 2011). ومن الذي يكون مسئولاً عن ضمان الخصوصية، المطورون أم الوكالات أم المستخدمون، وعن الوسائل التقنية والإدارية لتنفيذ هذه الأمور (Coterill 2011). وما الذي يشكل المعلومات الخاصة وكيفية تأطير الخصوصية حول المخاطر ومدى الضرر بدلاً من تأطيرها حول تعريف المحتوى (Hen





et al. 2011). واقــترح هؤلاء العلماء أيضًا دخول الأفراد في شراكات مع المطورين حيث يكــون بإمكانهم أن يختاروا بشــكل أكثر اســتباقية ما البيانات التــي يرغبون في نشرها، ولمــن تُنــشر هذه البيانات، وتحت أي ظروف تُنشر هذه البيانــات. كما اقترحوا أن تتيح الشركات للمســتخدمين الوصول إلى بياناتهم الشخصية بصيغة قابلة للاستخدام وذلك من أجــل مصلحتهم الشــخصية. وأخيراً، اقترحوا أن تقوم الشركات بمشــاركة الثروات الناتجة عن تحويل البيانات الشــخصية إلى أموال (Tene and Polonetsky 2012; Rubinstein 2013). وكمثال على هذه المشــاركة المفيدة لكلا الطرفــين في الثروات المتحصلة من بيع بيانات الشبكات الذكية حيث يتم استخدام البيانات التي تم تكوينها من خلال العدادات الذكية والمتعلقة باســتهلاك أفراد الأسرة للكهرباء من قبل شركة الكهرباء من أجل ضمان كفاءة التزويد بالكهرباء، ويتم مع ذلك توفير تطبيقات لأفراد الأسرة لتمكينهم من مراقبة استخدامهم الشخصي وأن يعدلوا سلوكياتهم من أجل توفير المال. إن سوق الصناعة، إلى حد كبير، يُريد إما الاستمرار في الإمدادات الحالية على ما هي عليه أو تقليل هذه الإمدادات، مع إدارة الخصوصية من خلال تنظيم يقود الســوق ولا يكبح الاســتفادة الاقتصادية من البيانات. وفي هذه الحالة، فمن المتصور أن الشركات ســوف تقوم بعملية التنظيم الذاتي خشية حدوث أضرار بسمعة تلك الشركات أو أن يعبر عملاء الشركات عن عدم رضاهم من خلال عدم التعامل مع هذه الشركات إذا انتهكت هذه الشركات خصوصية هؤلاء العملاء بشكل مبالغ فيه (Minelli et al. 2013; Mayer-Schonberger and Cukier 2013).

أمن البيانات (Data Security):

مع وضع قيمة البيانات في الاعتبار، وخاصة البيانات الشــخصية التي يمكنها أن تســهل سرقة الهوية، أو البيانات التجارية التي يمكن قرصنتها أو اســتخدامها من أجل تحقيق ميزة تنافســية، أصبح أمن البيانات مجالاً هامــاً في حماية البيانات. فالأفراد يمكنهم فتح العديد من الحســابات ســواءً الحسابات الشــخصية أو حســابات العمل، مع العديد من الهيئات والبائعين وذلك باســتخدام العديد من رموز الهوية الشخصية وأسماء المستخدمين وكلمات السر. فالبيانات المتعلقة بهؤلاء الأفراد، والبيانات الشــخصية الخاصة بهم، يتم تخزينها في الكثير من الأماكن (على ســبيل المثال أجهزة الحاســب الآلي، والأجهــزة الرقمية، والهواتف



الذكية، ومحركات الأقراص الصلبــة الخارجية، وأجهزة تخزين البيانات المتنقلة مثل (USB Drive) والخوادم. وتكون هذه الحسابات والأجهزة الرقمية عرضة للقرصنة (الدخول على الحســابات أو الوصــول إلى الأجهزة)، والبرمجيات الخبيثة (الرموز التي يتم تثبيتها بشــكل سري على جهاز الحاسب والتي يمكنها الوصول إلى المعلومات الحساسة أو تسجيل ضغطات لوحة المفاتيح وترســل هذه الأشــياء إلى طرف ثالــث)، والانتحال (Phishing) (الرســائل الإلكترونية الاحتيالية أو الهندســة الاجتماعية من خــلال المكالمات الهاتفية التي تحاول أن تجعل الشخص يتطوع بتقديم معلومات حسابه)، مما يؤدي إلى سرقة واختلاس البيانات.

وقد ناقش Gantz and Reinsel (2011) بالتفصيل خمســة مستويات لأمن البيانات، وكل مســتوى من هذه المستويات له محركات مختلفة إلى حد ما، وذلك على الرغم من أن جميع هذه المستويات تتطلب إجراءات أمن استباقية. وهذه المستويات تكون على النحو التالي:

(١) الخصوصية: من أجل المحافظة على المعلومات وتقييد تداولها.

(٢) الامتثال والالتزام: من أجل حماية البيانات التي من المحتمل أن تكون قابلة للكشــف عند التقاضي، أو الخاضعة لشروط الاستبقاء.

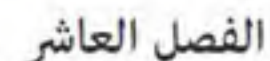
(٣) الاحتراز: من أجل حماية البيانات التي يمكنها أن تؤدي إلى أو تساعد على سرقة الهوية.

(٤) السرية: من أجل التأكد من حماية المعلومات السرية مثل الأسرار التجارية.

(٥) التأمــين، من أجل حمايــة المعلومات عالية السرية مثل الصفقــات التجارية، والملفات الشــخصية، والســجلات الطبية، والمعلومات العســكرية، التي يمكن أن يكون لها تأثير يتجاوز السرقة الشخصية.

وتُعد إدارة المستويات الخمسة السابقة لأمن البيانات مُهمة بالغة الأهمية بالنسبة للأفراد والشركات والمؤسسات. وعلى وجه العموم، يتم إنجاز هذه المُهمة من خلال التحكم في الوصول عبر برامج الأمن وتصميم النظام / الشــبكات (على ســبيل المثال تثبيت جــدار ناري، وبرامج لاكتشــاف ومكافحة الفيروســات، والرســائل غير المرغوب فيها، وحصان طروادة، والبرمجيات الخبيثة)، والتشــفير الذي يتطلب كلمة سر لفتح الملفات المغلقة. وإذا وضعنا في الاعتبار أنه





بمجرد أن يتم الوصول إلى البيانات فسوف يصبح من السهل نسخها ونشرها، فإن إدارة الحقوق الرقمية تسعى لتقييد مثل هذه الممارسات وأن تجعل من السهل أيضًا تتبع تلك الممارسات.

ومع ذلك، وعلى الرغم من هذه التهديدات، فإن الأجهزة الرقمية والخدمات والبيانات، ونقــاط الضعف المحتملة في تهيئتهم، تزيد بسرعة أكــبر من المقدرة على تأمينهم (Gantz and Reinsel 2011). وفي حين أن الكثير من أجهزة الحاســب الشخصية وأجهزة الحاسب المحمولــة آمنة بصورة نســبية مادامــت برامج الحماية الخاصة بها تُحدَّث بشــكل دوري، إلا أن أجهزة الحاســب اللوحية، والهواتف المحمولة، وأجهــزة الفاكس، ومحركات الأقراص الصلبة الخارجية، والأجهزة الطرفية (مثل الطابعات، وأجهزة المســح الضوئي، وآلات تصوير المستندات الشبكية)، والعدد الهائل من الأجهزة التي تشكل إنترنت الأشياء عادة ما تتمتع بحمايــة قليلــة (Rezendes and Stephenson 2013; Rose 2013). وقــد أدى ذلك بـ Mims (2013) إلى الادعاء بأنه بمجرد وجود أي شيء في بيئة تشــتمل على رمز رقمي ويتم توصيله بالشــبكة، فيمكن قرصنة هذا الــشيء من خلال كائنات يُحتمل أن يتم التحكم بها عن بعد، أو يتم إجبار ذلك الشيء على تجاوز معايير التصميم، أو العمل بطرق غير مريحة أو خطرة، أو يضلل المســتخدمين مما يؤدي إلى حدوث الأخطاء والأضرار، أو يتجســس على الســاكنين أو المســتخدمين. وعلاوةً على ذلك، فإن الطرق المستخدمة من قبل القراصنة أو المحتالين أصبحت أكثر تعقيداً وعدوانية عن الســابق، وأن صناعة الأمن دوماً في ســباق من أجل مواكبة هؤلاء القراصنة والمحتالين (Goldberg 2012).

فمــع تطور ثــورة البيانات، ووفرة المزيد مــن الأجهزة التي تُنتج، وتتبادل، وتســتخدم البيانــات، يبدو أن القضايا والمخاوف الأمنية مُعرضة للازدياد وليس التقلص (Gantz et al. 2007). وهذا بدوره ســيؤدي إلى تفاقم الجرائم مثــل سرقة الهوية، وتقوض الثقة في نظم البيانات، ورفع سلسلة من المســائل القانونية المتعلقة بالمسؤولية والالتزامات تجاه حماية النظم عندما يتم العبث بالبيانات، واختلاسها وسرقتها (Weber 2010).

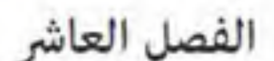


التشـخيص، والتصنيف الاجتماعـي والخطوط الحمـراء (Profiling, Social Sorting and Redlining):

بعيداً عن توسـيع محيط المراقبة، وانتهاك الخصوصيـة، ونقاط الضعف المتعلقة بأمن البيانات، تأتي المخاوف بشأن ماهية المجالات التي يتم فيها توظيف فيضان البيانات لتحقيق أشياء معينة. وكما أشرنا إلى ذلك في الفصل السابع، لم تتغير طبيعة البيانات فحسب، ولكن يوجـد أيضًـا تحول نوعي في الكيفيـة التي يتم بها تحليل البيانـات، وفي أي الغايات يمكن اسـتخدام البيانات وتحليلات البيانات، ومن لديه السلطة لاستخدام واستخلاص القيمة من البيانات. فالبيانات كانت تسـتخدم لوقت طويل من أجل تنميط وتقسيم وإدارة السكان، ولكن هـذه العمليات أصبحت في الوقـت الراهن أكثر تعقيداً، ودقة، وانتشـارًا، وانتظاماً عـن ذي قبل بكثـير. وفي حين أن الهيئات الحكومية تكون ملفات شـخصية للأفراد وذلك مـن أجل أغراض الأمـن ورصد الاحتيال، إلا أن معظم النمو في عملية التنميط يتم من قبل الشركات التجارية والتي تسـعى إلى فهم واسـتهداف قاعدة المستهلكين الحالية والمحتملة (انظر الفصل الثاني).

ففي الماضي، كانت الشركات تقوم بالتسـويق الجماعي، ونشر رسـالة عامة إلى حد ما وسط مجموعة كبيرة جداً من الجمهور وذلك من خلال الإعلان الجماعي باستخدام التلفاز، والراديـو، والطباعة، والبريد الإلكتروني، وذلك من أجل الوصول إلى المجموعة المسـتهدفة لتلـك الـشركات (Schwartz and Solove 2011). وبمرور الوقت، تـم تعزيز هذا الأمر من خلال التسويق المستهدف وذلك باستخدام التنميط الخام بشكل نسبي. وهنا، قامت مجموعة صغيرة من الشركات المتخصصة بتكوين تصنيفات عامة للسكان حيث تم تصنيف الأسر إلى أصناف ذات ملفات شخصية عامة، والتي كانت في العادة يتم تسميتها من خلال اختيار متغيرات ومواقع ديموغرافية، على سـبيل المثال باستخدام بيانات الإحصاء السكاني أو البيانات التي تعكس أسـلوب الحياة كاسـتخدام بيانات الاشتراكات في دوريات معينة على سـبيل المثال (Goss 1995). وبدلاً من محاولة تنميط عملائهم الشـخصيين، ترغب الشركات في شراء الملفات الشـخصية وبيانات الاتصال وذلك من أجل تقسـيم واستهداف أصناف معينة من السكان من خلال رسالة مؤطرة بدقة، وفي الوقت نفسه تحقيق الفعالية





في ميزانية التسـويق من خلال تقليـل مصروفات الإعلانات المبـددة. وفي الآونة الأخيرة، بدأت شركات التنميط في إنشـاء ملفات شخصية فردية بدلاً من الملفات الشخصية العامة وذلـك من خلال تجميـع البيانات من مصادر عـدة مثل معاملات البطاقـات الائتمانية ومعاملات بطاقات المتاجر، ومسـارات النقر، والمشـاركات في وسائل التواصل الاجتماعي، والعديد من أنواع البيانات الشـخصية الأخرى (Siegel 2013). وعلاوةً على ذلك، تستفيد الشركات ذاتها من الكميات الهائلة من البيانات التي تنشئها عن العملاء على سبيل المثال، مـن خلال المبيعات ومسـارات النقر وإدارة علاقات العملاء، وتجميـع هذه البيانات مع المصادر الأخرى للبيانات وذلك من أجل تكوين الملفات الشـخصية الخاصة بهم والاشتراك في التسويق السلوكي.

هنـا من الناحية الظاهرية، يبـدو أن التنميط التنبئي موقف مربـح للجميع، للعملاء والبائعين، فالعملاء يحصلون على معاملة شخصية ومخصصة، في حين يحصل البائعون على مبيعات أكثر ويقللون من النفقات. ولكن وفي المقابل، فإنه يمكن استخدام التنميط التنبئي من أجـل تصنيف الناس من الناحية الاجتماعية واسـتبعاد فئات منهـم من خلال اختيار أصنـاف معينة منهم للحصول على وضع تفضيلي وتهميش واسـتبعاد الآخرين (Graham 2005 ; Leyshon and Thrift 1999). فعلى سبيل المثال، يمكن للشركة تنميط الناس فيما يتعلق بمخاطرهم الائتمانية وما هي احتمالية مقدرتهم على الوفاء بسـداد مدفوعاتهم، أو فيما يتعلق بقيمة حياتهم المتوقعة إذا ما ظلوا في برنامج الولاء، وما هي احتمالية تغييرهم لعاداتهـم (Minelli et al. 2013). ومـن خلال اسـتخدام مثل هذه الملفات الشـخصية والتنميـط، يمكن للشركات تحديد الأولويات لاهتماماتهـم ومواردهم، وذلك يتم عادة من خلال التركيز على العملاء ذوي القيمة العالية. ومن ثم يتم قياس وتصنيف العملاء بشكل منتظم ويحصلون على خدمات تفضيلية، وبعضها يكون تمييزياً مثل استبعاد السكان الذين يُعَدُّون بلا فائدة (لن يحققوا مكاسب للشركة)، أو لن يحققوا مكاسب كافية، أو أن المخاطر المرتبطة بهم عالية، مما يمكن أن يؤثر بدوره على الفرص الحياتية (Curry 1997; Danna and Gandy 2002; Wyly in press). فهـذه الملفـات الشـخصية تُعتبر منتجات ويمكن بيعهـا لشركات أخرى تكون مهتمة في تقييم قيمة ومخاطر العملاء المحتملين، ومن ثم فإن النتائج السلبية يمكنها أن تطال الشخص عبر الأماكن وعبر الأزمنة.



ومثل هذه الممارسات التمييزية يمكن أن تشتمل على تسعير فعال وشخصي، فمن الأمور الشائعة بالفعل في سلاسل المتاجر أن تكون السلعة ذاتها مسعرة بطرق مختلفة عبر المتاجر المختلفة وذلك استناداً إلى خصائص وسمات السكان الذين يتسوقون من هذه المتاجر، أو أن يكون سعر المنتجات مختلفاً وفق الكمية المشتراة (مثلاً المنتج الواحد بدولار في حين أن الثلاثة منتجات بدولارين)، أو أن تختلف الأسعار باختلاف الجماعات (مثلاً حصول الطلاب أو المواطنين الكبار في السن على تخفيضات) (Varian 1996). وتتمثل رغبة الكثير من الشركات في تداول مثل هذه الممارسات على أسس فردية، ومعدلة لكي تناسب الملفات الشخصية، وذلك حتى يدفع أناس مختلفون أسعار مختلفة للمنتج ذاته (كما هي الحال مع أسعار تذاكر الطيران ولكن بناءً على نموذج مخصص). وتُراوح الأسعار أيضًا بشكل فعال ووفق السياق، بناء على الظروف. وقد بدا بالفعل تطبيق مثل هذه الممارسات الشخصية والمرنة، بما في ذلك قطاع المتاجر. فعلى سبيل المثال، تطرق Clifford (2012) إلى أمثلة تبين أن المتسوقين يدفعون في المتجر ذاته أسعار مختلفة للمنتج ذاته وذلك استناداً إلى ملف العميل الشخصي. وفي بعض الحالات، تقوم المتاجر بتعديل الأسعار وقت تسوق العملاء وذلك بناءً على موقعهم داخل المتجر وتاريخ التسوق لهؤلاء العملاء حيث يتم منح القسائم الإلكترونية في الموقع عندما يقوم المتسوقون بمسح المنتج باستخدام التطبيقات في هواتفهم الخلوية. إن الهدف من وراء التسعير الشخصي والفعال هو الحصول على كامل فائض القيمة المتاح في المعاملة وذلك من خلال تسعير البضائع والخدمات لأقرب سعر ممكن من السعر الاحتياطي (سعر التحفظ) لدى الفرد (Tene and Polonetsky 2012: 17)، وبعبارة أخرى، فإن هذا مصمم للاستفادة من الإنفاق المثالي لمصلحة المتجر.

وبالنظر إلى قدرتهم على تشكيل وتصنيف خبرات المستهلكين والفرص الحياتية، فإن الملفات الشخصية للأفراد والبيانات وممارسات البيانات التي تدعمها بحاجة إلى اهتمام أكبر. وفي الوقت الحالي، فإن كلاً من البيانات المستخدمة والتحليلات المطبقة مبهمة وغامضة بالنسبة للمستهلكين والمواطنين، فإذا انتهى الأمر بالشخص مدرجاً على قائمة الممنوعين، أو تم معاملته بشكل سلبي، أو تم استبعاده من منتج ما من قبل شركة ما، على سبيل المثال تم وضعه في القائمة السوداء للحصول على الائتمان، فمن الصعوبة بمكان تحديد الأسس لهذا التمييز ومجابهة هذه الأسباب. فالنقاد يدعون بأننا قد ولجنا عصراً حيث:



ربمـا يتراءى للبعض أن لبنية البيانات أو الصورة التخيلية للبيانات ميزة وجودية، ولكنها بالفعـل ذات ميزة وجودية. إن مـا تقوله بنية بياناتك عنك أكثر حقيقة وصدقاً مما تقوله أنـت عن نفسـك، فبنية البيانات هي البنية التي تم الحكـم عليك من خلالها في المجتمع، وهي البنية التي تملي وتفرض وضعك في العالم، فما نشاهده عند هذه الفترة من الزمن هو غلبة التمثيل على الكينونة (Critical Art Ensemble 1995).

وعـلى الرغم من ذلك فنحن نعرف أن البيانات الكبيرة وبنية البيانات فوضويان - فهما ممتلئان بالثغرات، والمتناقضات والأخطاء (انظر الفصل التاسـع) - ونعرف أيضًا أن البرامج والخوارزميـات تعمل بطرق انتقائية ومتحيـزة (Kitchin and Dodge 2011). ومن ثم، فـإن الأنظمة التي تعتمـد على البيانات الكبيرة، والتي لا تُعد شـفافة وتفتقد إلى التدقيق اللازم وحق الرجوع الملائم، تطرح جميع الأنواع المحتملة من المشاكل وذلك عندما يمنحون الأفضليـة لبنية البيانات. وفي الواقع، فـإن الوقت أصبح ملائماً الآن أكثر من أي وقت مضى للرقابة القانونية والتنظيمية على التنميط التنبئي، وعلى الإجراءات التي تشوش على التمييز الوجودي للبيانات وتحمي الناس من الممارسات التمييزية والضارة.

الاسـتخدامات الثانوية والتحكم في الانسلال والحوكمة الاستباقية (Secondary Uses, Control Creep and Anticipatory Governance).

إن أحد الأسس التي تستند إليها سياسة الخصوصية وحماية البيانات في الاتحاد الأوروبي وأمريـكا الشـمالية هو مفهوم تقليص البيانـات. وهذا المفهوم يشـترط أن تقوم الهيئات والبائعـون بإيجـاد البيانـات الضرورية فحسـب لأداء مهمة ما، بمعنـى أن يتم الاحتفاظ بالبيانـات طالما كان ذلك ضرورياً لأداء هذه المهمـة أو كما تمليه الاعتبارات القانونية، وأن البيانات التي يتم تكوينها يتعين اسـتخدامها فحسـب من أجل هـذه المهمة (Tene and Polonetsky 2012). وبمعنى آخر، لا يتعين على جامعي البيانات أن يستولوا على أي شيء يمكنهم الاسـتيلاء عليه، ولا يتعين عليهم كذلك تخزين هذه البيانات لمدة غير محدودة أو أن يتيحوا هذه البيانات لاستخدامات ثانوية. ومن الواضح أن هذه التطلعات والطموحات تتعارض مع المبادئ والأسس المنطقية للبيانات الكبيرة وعمل أسواق البيانات التي تسعى إلى تكديس البيانات في حال كان من الممكن أن تكتسـب هذه البيانات أي قيمة في المستقبل،



(Tene and Polonetsky 2012; Andrejevic 2013). فالحل أمام البائعين تمثل في إعادة تقديم البيانات من خلال إزالة الأجزاء التي تشير لهوية الأفراد من هذه البيانات أو إنشاء بيانات مشتقة، مع إخضاع مجموعة البيانات الأساسية فقط لعملية تقليص البيانات. ويمكن في هذه الحالة بيع البيانات المعاد تقديمها أو استخدامها بعدة طرق، ومثل هذا الأسلوب لإعادة تحديد الغرض من البيانات أمر شائع، حيث يجادل البائعون أنه لا توجد حاجة لإبلاغ الشخص الذي تشير إليه البيانات أو الحصول على موافقته، حيث إن هذه البيانات مجهولة المصدر، ومشتقة ومجمعة (Solove 2007). وهذا الحصاد المحتمل للكميات الهائلة من البيانات، والذي تم الحصول على معظمه دون معرفة أو فهم الأشخاص، والذي يتم استخدامه بعد ذلك في استخدامات ثانوية، يثير الكثير من الأسئلة الأخلاقية التي لا تتعلق بالخصوصية وحماية البيانات فحسب، بل تتعلق كذلك بالحوكمة، ومن الأمثلة الواضحة والمتعلقة بالحوكمة هو التحكم في الانسلال.

إن التحكم في الانسلال يحدث عندما تكون البيانات التي تم إنشاؤها لشكل من أشكال الحوكمة مشابه لشكل آخر (Innes 2001). وحدث هذا غالباً بشكل شديد الوضوح فيما يتعلق بالأمن وخصوصاً بعد أحداث الحادي عشر من سبتمبر، حيث تم إعادة توظيف البيانات الخاصة بخطوط الطيران والبيانات الإدارية الحكومية من أجل تنميط وتقييم مخاطر الأمن للركاب (Lyon 2003b). وعلى نحو مشابه، فلقد تم إعادة توظيف كاميرات الإشارات المرورية ورسوم الازدحام في لندن من أجل القيام بمهام تتعلق بالأمن، وذلك بدلاً من مراقبة الانتهاكات المرورية فحسب (Dodge and Kitchin 2007a). ومن الأمثلة التجارية على التحكم في الانسلال هو نظام الملاحة في السيارات المستأجرة حيث يتم إعادة توظيف هذا النظام من مساعدة السائقين على إيجاد طريقهم فحسب إلى مراقبة وفرض غرامة على هؤلاء الذين يقودون خارج المسار أو يخرجون من الولاية (Elliott 2004). فالتحكم في الانسلال يقوض بشكل منهجي الأسباب المنطقية وراء تقليص البيانات، في حين أن نشره يشكل تهديدات واضحة للحريات المدنية، حيث يكون جميع المواطنين، سواء الأبرياء أو المذنبين، خاضعين لمراقبته وإجراءاته التأديبية.



وهنـــاك مثال آخر على الاســـتخدام الثانوي للبيانات والذي ينتهـــك الحريات المدنية ألا وهـــو الحوكمة الاســـتباقية (التوقعية). وهنا، يتم اســـتخدام التحليـــلات التنبئية من أجل تقييم السلوكيات أو الأحداث المستقبلية المحتملة والتوجيه باتخاذ الإجراء المناسب. فمثل هذه الحوكمة الاســـتباقية تُعد ســـمة من ســـمات الســـفر بالجو خلال عدد من السنوات، حيث يتم تنميط المســـافرين من أجل المخاطر والتحقق الأمني وذلك قبل البدء في الرحلة (Dodge and Kitchin 2004). وبالمثل، تســـتخدم العديد من قوات الشرطة في الولايات المتحدة الأمريكية التحليلات التنبئية من أجل توقع موقع الجرائم المســـتقبلية أو من أجل توجيه ضباط الشرطة لزيادة دورياتهم في هذه المناطق (Siegel 2013). وعلى النحو ذاته، تســـتخدم بعض الشركات مثل هيوليت باكارد (Hewlett Packard) التحليلات التنبئية من أجل تقييم ومعرفة مَنْ مِنَ الموظفين من المحتمل أن يغادر الشركة، ومن ثم التدخل بشكل استباقي من أجل تقليل نفقات استبدال الموظفين (Siegel 2013). وفي مثل هذه الحالات، فإن آثار البيانات الخاصة بالشـــخص تقوم بما هو أكثر من تتبع الأشخاص، إذ تقوم باستباق هؤلاء الأشخاص (Stalder 2002).

إن مثل هذه الأســـاليب من الحوكمة تفرض قضايا أخلاقية بســـبب أنها توجه الاهتمام إلى مجموعات وأماكن معينة، وهي كذلك تســـعى لمراقبة الســـلوكيات التي ربما لن تحدث على الإطلاق، وخلال هذه العملية، كما تقوم بإعادة تشـــكيل كيف يمكن للناس العمل من خـــلال الانضباط الـــذاتي (Harcourt 2006). وعلاوةً على ذلك، فـــإن مثل هذه التحليلات التنبئيـــة تعمل في غالب الأحيان على تفاقم التحامل والتمييز، وتعمل كذلك على تأســـيس النبوءات المحققة لذاتها (التوجه نحو إشـــباع الـــذات) (Harcourt 2006; Siegel 2013). ومن ثم فلقد ادعى Raley (2013: 128): "إن البيانات في هذا المجال تكون أدائية: فتجميع أجزاء وبتات (bits) من البيانات في ملفات شخصية لمشتبه به في الإرهاب، وإعادة تأسيس البيانات المجردة عند استهداف حياة واقعية سوف يكون لهما تأثير على إنتاج هذه الحياة، هذا الشـــخص، كمشتبه به بالإرهاب بالفعل". ومن ثم يتم وضع الفرد تحت بؤرة المراقبة وممارسات مراقبة الدولة، وجعله خاضعاً لنتائج وتداعيات كونه مشتبهاً به بالإرهاب، على الرغـــم مـــن عدم التيقن من هذا الاشـــتباه. وبعبارة أخرى، تتم معاملة الشـــخص بطريقة مختلفـــة من جـــراء التوقع بأنه ربما يكون شـــيئاً مـــا، في حين أنه ربما يُكن هـــذا الشيء أو



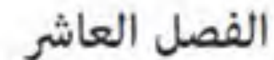
لا يكنــه أو ربمــا يرتكــب هذا الفعــل أو لا يرتكبه. وفي حــين أن النوايــا وراء هذا العمل ربمــا تكون نوايــا عظيمة وحســنة، إلا أن التوقع في هذه الحالة له تداعيــات تتجاوز منع الأحداث المتوقعة (Harcourt 2006). وكما هو الحال مع الأشــكال الأخرى من التنميط، فإن الحوكمة التنبئية يتم توجيهها بشكل كبير من خلال خوارزميات الصندوق الأسود التي تعمل على بيانات ذات أصل غير معروف، وهي على وجه العموم منغلقة على حق الرجوع (النقــد). فالحوكمــة التنبئية كانت مثار القليل للغاية من الاهتمام النقدي أو المناقشــات وذلــك بما يتعلق بتأثيراتها، وعلى الرغم من ذلك تم تطويرها بفعالية من أجل اســتخدامها في الحكومات والشركات.

أســاليب الحوكمــة والضوابــط التقنيــة (Modes of Governance and Technological Lock-Ins):

بعيداً عن التحكم في الانسلال والتنميط التنبئي، فإن ثورة البيانات لها تأثيرات محتملة فيــما يتعلق بتنظيم وتشــغيل الحوكمة. وإذا وضعنا في الاعتبار الشــبكة المتســعة لجمع البيانــات، وقراراتها الدقيقة، فــإن قلق الكثير من المعلقين ينبع من أن عصر ظهور البيانات الكبيرة يسبق عصر ظهور ما أصطلح على تسميته وصاية الأخ الأكبر (Big Brother). وهذا يعني أننا في عصر أضحت فيه الدولة والمؤسسات تعرف وتتوقع الكثير جداً عن الأفراد من خلال مراقبة البيانات والتنميط التنبئي لدرجة أنهم يمتلكون الســلطة لفرض أشكال قاسية وضارة من التحكم الانضباطي.

وكما أشرنا إلى ذلك خلال هذا الفصل، فنحن نعيش وبدون شك في عصر المراقبة الزائدة بشكل مبالغ فيه، حيث أصبحت المراقبة أكثر اتصالاً من أي وقت مضى، وهي مدعومة من خلال تزايد ثقافة السيطرة التي ترغب في الأمن، والاستقرار وتقليل المخاطر (Lyon 2007). وهنــا، يتم تشريع نــوع جديد من الحوكمة، وهــو نوع يؤول فيه تنظيــم جوانب معينة من الحياة اليومية إلى الأنظمــة التقنية المملوءة بالتوكيل الثانوي (Kitchin and Dodge 2011). فعلى ســبيل المثال، فإن معرفة، ومعالجة وإدارة الانتهاكات المرورية تتم بشــكل متزايد من خلال الأنظمة الآلية، والمسموح بها من خلال برنامج يقوم بمعالجة البيانات التي تتعلــق بأرقام اللوحات، والسرعة، وحق الوصول، وإحالة هــذه البيانات إلى قواعد بيانات





الملكية من أجل فرض غرامات أو خصم نقاط بشكل آلي (Dodge and Kitchin 2007a). وعلى النقيض من الأشـكال التقليدية من المراقبة التي تسـعى إلى تشجيع الانضباط الذاتي، فإن الأنظمة الآلية تقوم بإعادة تشكيل السلوك بشكل فاعل (Agre 1994; Kitchin and Dodge 2011). وفي مثل هذه الأنظمة، فإن آليات إنشـاء البيانات وتنظيمها تعتبر مجالاً مهماً للغاية من النظام، وذلك بدلاً من كونها جزءاً خارجاً عن النظام. فعلى سـبيل المثال، بدلاً من أن يقوم عمال صناديق المحاسـبة بإجراء الانضباط الذاتي لمعدل عملهم لأنهم ربما أو ربمـا لا يعرفون أنه تتـم مراقبتهم من خلال كاميرات المراقبـة التلفزيونية ذات الدائرة المغلقة، فإن عمل مسـح البضائع ومعالجة المدفوعات هي الوسـائل التي يتم من خلالها مراقبة معدل العمل بشـكل مسـتمر. وفي مثل هذه الأنظمة/ فإن المراقبة وآثار البيانات المرتبطة بهذه المراقبة أصبحت مستمرة، ومتغلغلة، ومواظبة، ومتفاعلة، مع سلوك الخاضع لها، ولكنها تقع خارج سيطرته (Cohen 2012).

إن البيانات التي تكونت بشكل آلي لا يستخدم جميعها من أجل تشريع الإدارة الآلية، ولكن هذا هو الاتجاه السائد بشكل متزايد، وخاصةّ للأنظمة عالية التنظيم. وعلاوة على ذلك، فإن الأسـاليب الجديدة من التحكم، إلى الآن، لا يتم السـماح بها بالطريقة الصارمة التـي توقعتها رؤية جـورج أورويـل (George Orwell) في روايتـه "الأخ الأكبر"، فهذه الأسـاليب عادية وروتينية ومغرية وتشـجع المواطنين على أن يشـاركوا بفعالية وبتلقائية في ممارسـاتها (Kitchin and Dodge 2011; Cohen 2012). ونتيجـةّ لذلك، فإنه وكما ذكـر Solovo (2007)، فـإن البيانات الكبيرة لا تتعلق بشـكل كبير بمبـاشرة رؤية جورج أورويل ولكنها ترتبط بشـكل أكبر برؤية كافكا (Kafka) كونها تطرح قضايا عن الإجراءات الواجب اتخاذها فيـما يتعلق باللامبالاة، الأخطاء، الانتهاكات، الإحباط، نقص الشـفافية، والاسـتعانة أكثر مما تطرحه عن التسلط والاستبداد. وعلى النقيض من ذلك، فبالنسبة لما قاله Boellstroff (2013)، فإن المجاز الأسـاسي في البيانـات الكبيرة والمراقبة ليس الرؤية الكلية، ولكن بالأحرى هو التسليم، فبدلاً من أن يكون الفرد خاضعاً على نحو صرف لرقابة انتهاكية، يُسلم الناس بحماس ببياناتهم من خلال وسائل التواصل الاجتماعي، أو من خلال الانضـمام في برامج بطاقات الولاء... إلخ. وبغض النظـر عن الصياغة التصورية للعملية، فمن الواضح أن هذه الأنظمة الجديدة لها تأثيرات فيما يتعلق بالسلطة والسيطرة، وتقوم



بتشكيل التجارب والفرص الحياتية بطريقة مختلفة، وأن دور البيانات ومراقبة البيانات في مثل هذه الأنظمة يتطلب اهتماماً نقدياً إضافياً.

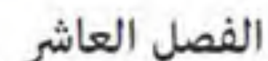
إن من الطرق التي يتم من خلالها تحويل الحوكمة من قبل التقنيات التي تدعمها البيانات أن نجعل الحوكمة أكثر تكنوقراطية بطبيعتها (معتمدة على التقنية بشكل كلي). فعلى سبيل المثال، فإن النزعة تجاه إدارة وتنظيم المدن من خلال أنظمة المعلومات والأنظمة التحليلية تعزز الأسلوب التكنوقراطي للحوكمة المدنية التي تفترض أنه من الممكن قياس ومراقبة جميع المجالات المدينة ومعاملة هذه المجالات على أنها مشاكل فنية يُمكن التعامل معها ومجابهتها من خلال الحلول الفنية (Kitchin 2014)، وهي توضح ما أطلق عليه (2013) Mattern مصطلح "العقلانية الأدائية"، وما أطلق عليه (Morozov 2013) "مقاربة الحلول التقنية" حيث يمكن تجزئة المواقف الاجتماعية المعقدة إلى مشاكل محددة بشكل دقيق يُمكن حلها أو البحث عن أفضل الحلول لها من خلال الحوسبة. ومرةً أخرى، فإن هذا يعد تجسيداً للبيانات الكبيرة، حيث يمكنها أن توفر إجابات لجميع المشاكل (Mattern 2013). ومن خلال التقاط وتمثيل الظواهر كبيانات بشكل آني أو لحظي، يصبح من الممكن من الناحية الظاهرية تخطيط، وفهم، وإدارة، وإصلاح الموقف أو المشكلة حال حدوثها. وقد أوضح (2013) Hill، أن "تفكير المدينة الذكية" ينم عن وجهة نظر تكنوقراطية بأن المدينة هي شيء يتعين علينا فهمه بالتفصيل كالمحرك أو محطة الطاقة النووية وذلك في حال كان لدينا بيانات كافية فقط. وأنه تبعاً لذلك سيمكننا السيطرة على هذه المدينة من خلال القوة الغاشمة للعلم والهندسة. وفي الواقع، فقد اقترح (2013) Matten أن تحضر البيانات الكبيرة يعاني من تراكم البيانات، وهو افتراض أن جميع التدفقات والأنشطة الهادفة يمكن الإحساس بها وقياسها. وداخل مثل هذا التفكير هناك افتراض واضح في الغالب وهو أن الكون تشكل من قبل المعايير القابلة للمعرفة والقابلة للتحديد والتي تؤكد لنا أنه لو كان بإمكاننا قياس جميع هذه المعايير، فسوف نكون قادرين على التنبؤ والاستجابة بتمكن تام وفقاً لذلك (Haque 2012). وعلى الرغم من ذلك، لن يكون بمقدور الحلول التقنية بذاتها حل المشاكل الهيكلية المتأصلة في المدن لأن هذه الحلول لا تتعامل مع أسباب تأصل تلك المشاكل. ولكن بدلاً من ذلك، فإن هذه الحلول تمكن فقط من إدارة مظاهر تلك المشاكل بشكل أكثر فعالية. وفي حين أن تقنيات المدن الذكية مثل



أســاليب التحليل اللحظية والآنية يتم الدفع بها كوصفة للتعامل مع جميع قضايا الحوكمة المدنيــة، إلا أن هذه التقنيات في الغالب الأعم تداري الشــقوق بدلاً مــن إصلاحها، إلا إذا واكب هذه التقنيات مجموعة أخرى من السياســات (Kitchin 2014). ويمكن أن ينطبق الأمر نفسه على الطرق التكنوقراطية المطبقة في بعض المجالات الأخرى، فأساليب الحوكمة التي تقودها البيانات تضمن بالفعل تنفيذ إســتراتيجيات مدعومة بالأسانيد ولكن يجب أن تكون مقترنة بالســياق، ومرنة، ومستهدفة بشكل ملائم، ولا يتم إدارتها ببساطة من خلال الأرقام والخوارزميات.

ويتزامن مع ذلك الانتقاد القائل أن الحوكمة أصبحت مبنية على معرفة تامة وتكنوقراطية مفرطة ومع الخوف من أن يتم الاســتيلاء عليها وتشــكيلها بشــكل صريح من قبل مصالح الشركات من أجل تحقيق مكاســب شــخصية. وبالعودة إلى المثال المتعلق ببرنامج المدينة الذكيــة ومثــال الحوكمة البلدية، وكما أشرنا إلى ذلك في الفصل الســابع، فإن عدداً من أكبر شركات خدمات البرامج والأجهزة يســعى بشكل فعال إلى التعاقد من الباطن على مجالات مــن الإدارة المدنية وأن تجعــل من بضاعتهم جزءًا لا يتجــزأ وذات دور جوهري في كيفية مراقبة وتنظيم العديد من مجالات الحياة في المدينة (Kitchin 2014; Townsend 2013). إن الهاجس والقلق النابع من خصخصة الإدارة المدنية ذو ثلاثة أوجه (Kitchin 2014):

- أولاً: أنه يعزز بشكل فعال الاقتصاد السياسي لليبرالية الجديدة والتحول نحو اقتصاد الســوق في القطــاع العام حيث يتــم إدارة وظائف المدينة من أجــل تحقيق الربح الخاص (Hollands 2008).
- ثانياً: أنه يؤسس للقيد التقني الذي يجعل المدن مدينة وممنونة بالفضل لمنصات ومزودي خدمات تقنية معينة خلال مدة زمنية طويلة، مما يخلق مواقف احتكارية (Hill 2013). والخطر في هذه الحالة يكمن في تأســيس تبعية للمســار المؤسسي مما لا يمكن إلغاؤه أو التحول عنه بسهولة (Bates 2012).
- ثالثاً: أنه يؤدي إلى حلول مغلفة أساســها "نظام واحد مناســب لجميع المدن الذكية" والتي لا تضع في اعتبارها إلا بشــكل ضئيل للغاية خصوصيــة وتفرد الأماكن، والناس، والثقافات، وتقيد إدارة المدينة على أسلوب تكنوقراطي ذي رؤية ضيقة للحوكمة (Townsend 2013).





وإذا ما أستحضرنا الواقع، فإن شركة IBM في الوقت الحالي منتجاً يسمى "مركز عمليات IBM الـذكي" الذي يجمع العديد من الأنظمــة التي تم تصميمها لمدينة ريو دي جانيرو في منتج واحد يمكن تطبيقه على أي مدينة (Singer 2012c، انظر الفصل السابع). ولا يعني هــذا أن نقول إن القيود التي تفرضها الشركات أمر حتمي، ولكن من الواضح أن ذلك يُمثل رغبة عدد من أكبر الشركات العاملة في هذا المجال. وبالمثل، يتم اسـتهداف أشـكال أخرى من الحوكمة بشكل متزايد من خلال الشركات التي تسعى إلى إيجاد أسواق للحلول التقنية. ومرة أخرى، فإن تداعيات مثل هذه الأشكال من الحوكمة تتطلب الكثير من التدقيق.

الخاتمة:

ناقش هذا الفصل التداعيات الأخلاقية، والاجتماعية، والسياسـية، والقانونية للتغييرات التي تحدث مــن جراء تكوين البيانات ودمج البيانات واسـتخدام البيانات. ومن الواضح أنه في حين أن مثل هذه الممارسـات لها فوائد بالنسـبة للحكومـات والشركات والمواطنين، إلا أنها في الوقت نفســه لها تداعيات سـلبية وتمييزية. وإذا وضعنا في الاعتبار السرعة التي يتغير بها مشــهد البيانات، فإن مواكبة التطـورات وتحديد الآثار المحتملة لهذه التطورات، ودراســة ردود الفعل الاجتماعية والقانونية المناسـبة يعتبر تحديًا هائلاً. وفي الواقع، هناك الكثير من الأسـئلة الأساسـية المعيارية التي تتطلب دراسـة تأملية والتي تتعلق بمن يمكنه تكوين مجموعات البيانات والوصول إليها ومشـاركتها وتحليلها، ولأي غرض وفي أي سـياق وتحت أي معوقات أو قيود (boyd and Crawford 2012: 673). إن الحاجة إلى إجابات لمثـل هذه الأسـئلة أمر ملـح، وذلك إذا وضعنـا في الاعتبار المدى والتأثـيرات المترتبة على مراقبــة البيانات وانتهاك الخصوصية والأمن غـير المحكم للبيانات، والتأثيرات الضارة للفرز الاجتماعي والاسـتبعاد، والطبيعة والأسـاليب المتغيرة للحوكمة. فحتى لو كانت الاستجابة سريعة ومبالغ فيها، فإن الإجابات المترتبة سـتكون متنوعــة، حيث إن العديد من أصحاب المصلحة يسعون وراء إستراتيجيات بديلة من أجل التعامل مع المخاوف العديدة، بحسب ما يتضح من خلال المناهج المقترحة لتفتت وانتهاكات الخصوصية. ونتيجةً لذلك، يمكن للمرء أن يتوقع الكثير من المناقشـات في السـنوات القادمة فيما يتعلق بإدارة العناصر العديدة في ثورة البيانات. وعلى وجه الخصوص، وكما ناقشـنا في الفصل الثاني، فإنه من المحتمل أن

يوضع عمل سماسرة البيانات تحت التدقيق المتزايد سواء من جانب الحكومة أو المواطنين.

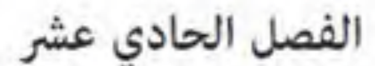
ومن منظور أكاديمي، من المهم للغاية أن يضطلع العلماء بدور فاعل في استقصاء ودراسة القضايا القانونية والسياسـية والاجتماعية والأخلاقية التي تنشـأ نتيجة لثورة البيانات، مع التتبع والإسهام في المناقشات العديدة المستمرة. يحتاج مثل هذا العمل إلى أن يشتمل من جانب على دراسات حالة تجريبية مفصلة توثق التأثيرات المادية والاستطرادية الناتجة عن كيفية توظيف البيانات، ومن جانب آخر على البحوث والرسـائل المعيارية والشـاملة التي تحـدد تأثيرات تجميع العديد من البيانـات والطرق البديلة الممكنة. وحتى هذه اللحظة، فإن معظم الإسـهامات كهذا الفصل والأعمال التي يناقشـها تكونت من الجانب الأخير مع وجود القليل جداً من الدراسـات التجريبية على بعض الأنظمة الاجتماعية الفنية المحددة ونتائج هذه الدراسات. إن مثل هذه الدراسات مهمة للغاية لأنها تكشف الطرق الملموسة التي تم من خلالها تقديم مبررات لتجميع البيانات، وكذلك المهام التي يتم ممارستها وكيف يتـم قبولها، وإعادة تنقيحها، ومقاومتها، وإفسـادها، وانتهاكها من قبل المجتمعات، وهذه الدراسـات توفر أيضًا قاعدة أدلة يتم من خلالهـا تصور كيف يمكن إعادة تأطير مثل هذه الأنظمة أو إعادة تحديد الغرض منها.



الفصل الحادي عشر: فهم ثورة البيانات

(Making Sense of the Data Revolution)

ناقش هذا الكتاب قضية أنه يتعين أن تكون هناك حاجة إلى تعاطٍ أكثر نقديةً وفلسفيةً مـع البيانات وأن ثورة البيانات ما زالت قائمة في الوقت الراهن. وفيما يتعلق بالأمر الأول، ادعـى البعـض أن البيانات تُمثل عناصر أساسـية لإنتاج المعرفة، ومع ذلـك، فحتى الوقت الراهن لم تحظ الصياغة الوجودية للبيانات والمجاميع المحيطة بإنتاج البيانات واستخدامها بالاهتـمام الكافي. وبدلاً من ذلـك، فإن التركيز كان منهجياً بصورة أكبر فيما يتعلق بتكوين البيانات، أو أن التركيز كان على الأشـكال الاشـتقاقية للمعلومات والمعارف. وفيما يتعلق بالأمـر الثاني، ادعى البعض أن تـدرج البيانات الصغيرة في البنى التحتيـة للبيانات وإتاحة الوصـول للبيانـات التي كان الوصول إليها محدوداً في السـابق، وظاهـرة البيانات الكبيرة تدمـر الابتكارات ولا تدعمها أو تحافظ عليها، فهي تغـير طبيعة البيانات من حيث حجم هـذه البيانات، وسرعتها، وشـموليتها، وعلائقيتها ودلاليتها، ومرونتها وتوسـعها، وتنوعها، وانفتاحها وتبادليتها، وهي السـبب في ظهور مجاميـع بيانات جديدة وطرق جديدة لفهم العـالم الذي نعيـش فيه. إن التأثـيرات التدميرية لثورة البيانات لهـا تداعيات هائلة على الحكومات والشركات والمجتمع المدني وذلك من خلال إنشـائها لمعارف وممارسات جديدة يمكـن اسـتخدامها من أجل إعادة صياغة كيف يمكن أن يتـم حكم الناس، وكيف يمكن أن تتم إدارة المنظمات، وكيف يمكن أن تتم زيادة القيمة والفائدة، وكيف يمكن أن يتم تكوين رأس المال، وكيف يمكن أن يتم تحسين الأماكن، وكيف يمكن أن تتم إعادة تخطيط وممارسة العلم. وهي أيضًا أثارت العديد من التحديات وطرحت الكثير من الأسئلة، ومكنت لوسائل أكثر شـيوعاً فيما يتعلق بمراقبة البيانات، وتقويض الخصوصية، وتنميط الأشخاص والأماكن والفرز الاجتماعي، ومهدت لسـن أشـكال جديدة من الحوكمة. فهذا الفصل الأخير يحدد خطة طريق دلالية من أجل إدراك وفهم البيانات وثورة البيانات وذلك بالنظر إلى الفجوات الحاليـة في الفكر المفاهيمي والمعرفة. ويشـير هذا الفصـل إلى أن مثل هذا الفهم بحاجة إلى أن يحـدث بطريقتـين: الأولى، من خلال التفكير الفلسـفي الشـامل والتحليل النقدي

والمفاهيمـي. والثانية، من خلال البحوث التجريبية المفصلة التي تتعلق بأصل، وتشـكيل، ووظيفية وتطور مجاميع البيانات.

فهم البيانات وثورة البيانات (Understanding Data and the Data Revolution):

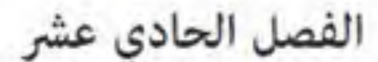
إذا وضعنا في الاعتبار فائدة وقيمة البيانات، فهناك حاجة ملحة للنظر إليهم من وجهة النظر الفلسفية والمفاهيمية. فالفلسفة، كما ادعى Wittgenstein (1921)، هي التوضيح المنطقـي للفكر، فكل مدرسـة من مدارس الفكر تُعَدُّ نظاماً للأفـكار ومنظوراً لفهم العالم. وبمصطلح عام، تتشـكل كل مدرسة وتختلف عن المدارس الأخرى من خلال أربعة مكونات -الوجوديـة (Ontology)، والمعرفيـة (Epistemology)، والأيدلوجيـة (Ideology)، والمنهجيـة (Methodology) - وهذه المكونات هي التي تحدد معاييرها. فالوجودية هي مجموعـة من المعتقدات التي تتعلق بطبيعة الوجود، وهي تهتم بما هو موجود وما يمكن ملاحظتـه ومعرفته. وبالنسـبة للبيانات، فالوجودية تتعلق بماهيـة البيانات. ما الطبيعة الأساسـية للبيانات؟ أما المعرفية فهي تتعلق بكيفية اسـتخلاص المعرفـة أو الوصول إليها، وهي الافتراضات عن كيف يمكننا معرفة العالم بمعنى كيف أصبح بإمكاننا معرفة البيانات؟ وهي أيضًا تتعلق بكيفية حشـد البيانات في ممارسـة التعلم. أمـا الأيدلوجية فهي تتعلق بالسياسـات الضمنية والغرض من إنتاج المعرفة، والمدى الذي يسعى من خلاله المنهج لكي يكـون محايداً بمعنى أن يكـون موضوعيًا وغير تقويمي، أو المدى الذي يسـعى من خلاله المنهج بفعالية للتدخل وتغيير العالم. وحيث ذلك، ينطوي تحت مفهوم الأيدولوجية العديد من الأسـئلة مثل هل البيانات سـابقة للحقيقة؟ وهل هي سابقة للتحليل وبلاغية؟ أو هل البيانـات اجتماعية أم سياسـية؟ وهل تمـت صياغة هذه البيانات بفعاليـة؟ أما المنهجية فهـي مجموعة الإجراءات المسـتخدمة من أجل تفعيل سـؤال أو نظرية ضمن الافتراضات الوجودية أو المعرفية للفلسـفة الإجمالية للبحث. وهي تشـكل اختيار الأساليب التي يتم من خلالها تكوين وتحليل البيانات، فالمنهجية والأسـاليب مرتبطان بشـكل وثيق، ولكن ما طبيعة هذه العلاقة بينهما؟

لقد شرع الفصل الأول في تقديم إجابات أولية عن هذه الأسئلة وقدم أيضًا صياغة لفهم البيانات من منظور فلسـفي مكن مـن فصل هذه البيانات عن كونهـا مجرد "مواد خام"



للمعلومــات والمعارف. فالبعض ادعــى أن البيانات اجتماعية ومادية، وأن البيانات لا تمثل العالم فحسب، ولكنها تنتج هذا العالم بشكل فعال. وادعى البعض أيضًا أن البيانات لا توجد بشكل مستقل عن الأفكار، والأســاليب، والتقنيات، والناس، والسياقات التي تنتج، وتعالج، وتديــر، وتحلل، وتخزن هــذه البيانات. وفي الواقع، يتم تنظيم البيانات وتخزينها في قواعد البيانات والبنى التحتية للبيانات التي تشكل لب المجاميع الاجتماعية الفنية المعقدة. وكما تم الإشارة إليه بالتفصيل في الفصل الأول، تشتمل هذه الأنظمة على أنظمة الفكر، وأشكال المعرفة، والتمويل، والاقتصاد الســياسي، وأنماط الحاكميــة، والتشريعات، والماديات، والبنى التحتية، والممارســات، والمنظمات والمؤسسات، والذاتيات والمجتمعات، والأماكن والأسواق، وكل نظــام من هذه الأنظمة يشــتمل على عنــاصر عديدة (انظر الجــدول ٣ - ١). فهذه العناصر تعمل سوياً بطرق متعددة، واتفاقية وارتباطية وذلك من أجل الإنتاج الاستطرادي والمادي لمجاميع البيانات (انظر الشكل ٣ - ١). لقد تم عبر هذا الكتاب مناقشة العديد من هذه المجاميع مثل البنى التحتية البحثية، والمحفوظات الوطنية، والبنى التحتية السيبرانية، ومشروعــات البيانات المفتوحة، وســماسرة البيانات، وأنظمة البيانــات الحكومية، وأنظمة الأعمال. وكل واحدة من هذه المجاميع، بالرغم من أنها تتشارك في قواسم مشتركة، إلا أنها تتطور وتعمل بطرق مختلفة، وهي تتكون من مجموعة من الأنظمة والعناصر التي تتوسع بشــكل مختلف (على ســبيل المثال من المنظمات، والماديات المحليــة، إلى الفرق المنتشرة، والقوانين الوطنية، وفوق الوطنية إلى الأســواق العالمية) ولكنها على الرغم من ذلك مرتبطة داخل كوكبة فريدة.

إن مثل هذه الصياغة تمكننا من جهة من التفكير في إنتاج البيانات وكيف يمكن صياغة هذه البيانات بأشــكال متعددة من النواحي الفنية، والسياســية، والاقتصادية، والأخلاقية، والمكانية، والزمانية، وصياغة اســتخدامات هذه البيانات والعمــل الذي تقوم به في العالم، وتمكننــا من الجهة الأخرى، من توجيــه البحوث التجريبية التي تتعلــق بمجاميع البيانات (انظر الجزء التالي). وهي، على الرغم من ذلك، مجرد مجموعة واحدة من الطرق المحتملة لفهــم البيانات، ومما لا شــك فيه أنه بمرور الوقت ســيتمكن العلماء مــن إنتاج مجموعة مــن المناظير التي يمكننا من خلالها فهم البيانات. إن تنوع وجهات النظر ســوف يؤســس لعناصر تكميلية مثمرة للأفكار الجديدة ووجهات النظر المفاهيمية من أجل توجيه البحوث



التجريبية. وربما تشـتمل هذه على تنظير البيانات من خلال منظار تأسـيسي (بنيوي) أكبر يركز على دور هذه البيانات في تشـغيل الرأسمالية المتأخرة، أو يستفيد من مفاهيم ما بعد البنيوية لـدولوز (أحد أهم الفلاسـفة الفرنسيين في النصف الثاني من القرن العشرين) التي تتعلق بأسـاليب الحوكمة الجذمورية (الجذمور: مفاهيم تتوزع على شـبكة، بحيث تلتقي مع بعضها دون اشتراط اللزوم)، أو الانتقادات النسائية ما بعد الاستعمار للنوع الاجتماعي والإنتاج المُسَيّس وتوظيف البيانات. وبغض النظر عن المنظور، فالمطلوب هو تفكير عميق، وحريص ونقدي مع تنظير العمل من خلال دراسات الحالة التجريبية.

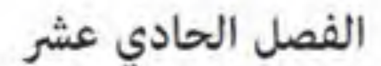
إننا لسـنا بحاجة إلى استكشـاف الصياغات المفاهيمية للبيانات فحسب، ولكننا بحاجة، كما ناقشـنا في الفصل الثامن، إلى دراسـة كيف تطرح ثورة البيانـات بعض التحديات أمام الفلسـفات الحالية للعلوم. فالبيانات الكبيرة وتحليلات البيانات توفر، على وجه الخصوص، إمكانيـة ظهـور نموذج علمي جديد، أي نهـج معرفي جديد يتميز بكثافة البيانات ويتسـم بالاستكشـافية بطبيعته وذلك من أجل تطوير النظرية العلمية. فالمحاولات الأولية للتفكير فيـما يمكن أن يعنيه مثل هذا التحول النموذجي كانـت في غالب الأحيان مرتبكة ومضللة، وتُعزز شكلاً من أشكال التجريبية التي يجدها الكثير من العلماء إبعادية وخطوة إلى الخلف وليسـت خطوة إلى الأمام. ويتعين اسـتبدال هـذا بمنظور أكثر فعالية لمـا يمكن أن يتكون منه العلم المسـاق بالبيانات من حيث معتقداته الفلسـفية والمنهجيـة. لقد بدأ مثل هذا التفكير بالكاد، ومع ذلك فهناك حاجة ملحة إليه من أجل مواكبة إيقاع التغير التقني ونشر المنهجيات المخصصة والبراجماتية (الواقعية)، وكذلك من أجل اسـتبدال الأشـكال التجريبية الضعيفة المنتشرة. فالتحول النموذجي أقل احتمالية في العلوم الإنسانية والاجتماعية، ولكن على الرغم من البيانات الكبيرة، فإن البيانات المفتوحة والبنى التحتية للبيانات توفر منهجيات وطرق جديدة للقضايا الاجتماعية، والثقافية، والسياسية، والاقتصادية، والتاريخية. فالفصل الثامن اسـتعرض خطة تفصيليـة بالمزايا والانتقادات النسـبية المتعلقة بالعلوم الإنسـانية الرقمية والعلوم الاجتماعية الحسابية التي يتم تطويرها، ولكن هناك حاجة ملحة لمناقشات كاملة وصريحة من شأنها استخراج النتائج من التحول المعرفي في الفكر والممارسة، ويجب أن تقوم مثل هذه المناقشات أيضًا بتقييم دور ومزايا دراسات البيانات الصغيرة، وكذلك المناهج الهجينة التي تمزج أساليب ووسائل البيانات الصغيرة والبيانات الكبيرة.



ومثـل هذا التفكير يجب أن يواكبه تفكير آخر، موجه بشـكل أكـبر تجاه المعيارية، في تكوين واسـتخدامات البيانات، وإنشاء وتشـغيل مجاميع البيانات. وكما ناقشنا في الفصل العاشر، تطرح ثورة البيانات عدداً من القضايا الشـائكة وذلك فيما يتعلق بالمدى الذي يتم على أساسه التقاط الحياة اليومية في مجاميع البيانات، وكيف أن بنية البيانات للناس تفوق وتضع قيوداً على حياتهم بجميع أنواع الطرق غير المتوقعة. فعلى سـبيل المثال، هل يجب أن يتحـلى الناس بالمقدرة على الوصول إلى جميـع البيانات المتعلقة بهم، وأن يكون لديهم الحـق في تصحيح والاعتراض على هـذه البيانات، أو تحليلات البيانات أو التنميط؟ وإلى أي مدى يجب أن يكون الناس قادرين على ممارسة الحق في أن يتم نسيانهم؟ وما هي الطرق التـي يمكن من خلالها إخضاع سـماسرة البيانات لرقابة وتنظيم مسـتقلين، وخاصة عندما تكـون بياناتهم وخوارزمياتهم تؤثر في الفرص الحياتية للنـاس؟ وما هو مدى تأمين الأنواع المختلفـة من البيانـات ؟ وما العقوبات المحتملة على انتهاكات هـذا التأمين؟ وما الحدود المشروعة للحكومات في تجميع ودمج البيانات عن مواطني هذه الحكومات؟ وإلى أي مدى يمكن إشراك المؤسسات في أعمال الحكومات وأنظمة الحوكمة؟ فهذه القضايا تتعلق بماهية المجتمـع المرغـوب فيه في عصر وفـرة البيانات، والإجابات عن هذه الأسـئلة لها تداعيات وخاصـةً فيما يتعلق بإمكانية إعـادة تصور وإعادة تهيئة مجاميع البيانات، وهذا يشـمل الأشـكال الجديدة من الحوكمة والتنظيم. فالعلماء القانونيون بدئوا بالفعل في التفكير في بعـض القضايا مثل نتائج خصوصية البيانات الكبيرة، وبـدؤوا كذلك في تقديم طرق بديلة للتعامـل مـع أكثر الآثار ضرراً، ولكن من الواضح أن مثل هذه القضايا بعيدة كل البعد عن الحل، وأنها ستكون محل تركيز المناقشات المتطورة والصراعات التشريعية، وأن هناك حاجة ملحة للإسهام في هذا الحوار الفعال والصريح والمستمر.

ويتزامن مع تطوير فهـم مفاهيمي ومعياري أعمق وأفضل للبيانات، الحاجة إلى توفير وجهات نظر شـاملة نقدية وكلية عن البنى التحتية للبيانات والبيانات المفتوحة والبيانات الكبيرة وتداعياتهم. لقد قدم هذا الكتاب تحليلاً شـاملاً بدلاً من طرح فرضية تسـتند إلى بحث تجريبي شامل، واستفاد من الكتابات الأكاديمية والإعلامية والصناعية. فهذا النهج ذو فائدة لأنه يجمع الأفكار والملاحظات سوياً من مصادر متعددة وذلك من أجل تقديم نظرة أعم وأشـمل عن المشهد المنتشر، مما يساعد على رسم وتصور الديناميكيات الأكبر، ودراسة





التأثـيرات المحتملــة لهــا. ومع ذلك، وعلى الرغم مما ســبق، فإن المنظــور العام الصعب، والتأملي، والشــامل لثورة البيانات لا يزال ضعيفاً في الواقع العملي. وبدلاً من ذلك، فهناك مجموعة من الكتــب وعلى وجه الخصوص في مجالات الإدارة، والأعمال، والمجالات التقنية تعمل بشــكل كبير على دعــم البيانات ولكنها تفتقر للمنظــور النقدي. وعلى النقيض من ذلك، فالمطلوب تقارير متعددة الاختصاصات ومعقدة في تفكيرها. وفي حين أنه من المفيد دراســة ثورة البيانات من خلال منظور متعدد الاختصاصــات والأوجه، إلا أن هذا المنظور مقيد بشكل جذري وذلك بالنظر إلى الأجندة الشاملة والآثار المترتبة. وعلى نحو مشابه، فإن المفهوم القائل بأن ثورة البيانات تنتج البيانات التي يمكن تفسيرها خارج سياقها أو المعارف خارج مجالها المخصص بحاجة إلى أن يتم إلغاؤه. وبدلاً من ذلك، فمن الضروري الاستفادة من الأفكار والمدارك عبر الإختصاصات، حيث يعمل العلماء الذين يتمتعون بخبرات مختلفة سوياً من أجل تقديم صورة أكثر شمولاً وعموميةً (Ruppert 2012).

وعــلاوةً على ما ســبق، فإن التحليلات أوجدت حاجة لكي نكــون حريصين ألا ننزلق إلى الجدليات التافهة التي إما تحدد بشــكل انفعالي حالة البيانات الكبيرة، والبيانات المفتوحة، والبنى التحتية للبيانات، أو توفر نقداً مســتحكماً يشــجب أكثر نتائجها سلبيةً. فالموقف لا يكون إما أبيضاً أو أسوداً، أو جيداً أو سيئاً، إلا فيما ندر، وعلى النحو الذي تطرحه الجدليات الانفعاليــة. وبدلاً من ذلك، يتعين علينا أن نقــر بأن التطورات الحادثة اتفاقية، وارتباطية، وســياقية، وهــي تنتشر بطرق فوضويــة وأحياناً تناقضية وتباينية. وكــما أشرنا إلى ذلك في الفصل التاســع، تعمل مجاميــع البيانات في الغالب على كلٍ من الإجبار والتحرير بشــكل فوري ومتزامن، وأن مســتهلكي البيانات أحياناً يكونون هــم منتجي البيانات، وأن الأنظمة تســعى إلى استنطاق المســتخدمين بطرق عديدة. وعلى هذا النحو، فإن المجاميع غالباً ما تكون متشابكة ويتعين أن يتم فتحها وتفريغها بشكل حذر وذلك من أجل إظهار تدرجاتها الرماديــة المتعددة. وعندما توجد تأثيرات تعمــل بطرق مضرة وتمييزية، فنحن بحاجة إلى تحديد الاستجابات الاجتماعية والقانونية المناسبة التي تحمي الحقوق الفردية والجماعية، ولكنها لا ترفض الكل لعدم رضاها عن الجزء.

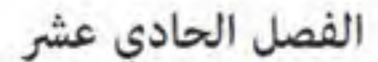


دراسة مجاميع البيانات (Researching Data Assemblages):

هنــاك ضرورة ملحة لإجراء بحــوث تجريبية مفصلة عن تكون، وعمل، وتعزيز مجاميع البيانات وذلك من أجل مواكبة وتعزيز التحليلات النقدية والمفاهيمية الأعم والأشمل. ففي الوقــت الحالي، لدينا فهم متعمق قليل بــكلٍ من البناء الإجمالي لمجاميع البيانات وأجهزتها وعناصرها الفردية. ونتيجةً لذلك، هناك حاجةً ملحة لإجراء دراســات حالة تتتبع الترتيبات الفنية الاجتماعية لكافة المجاميع، مثل تجميعة بيانات لسمسار ما أو البنية التحتية لبيانات بحــث ما، أو حركة البيانات المفتوحة في مدينة مــا، أو التوثيق المفصل لمجالات معينة من أحد المجاميع، مثل علم الاجتماع، والاقتصاد الســياسي لمطوري البرمجيات (Hackathons)، ومنتديات العطل الأسبوعية لمجموعات الاختصاص (Data Dives)، أو مجتمعات الممارسة داخل قطاع ما في البيانات الكبيرة. وبشــكل نموذجي، فإن مثل هذه الدراســات ســتكون دراسات مقارنة بطبيعتها، تقارن التكرارات في تجميعة ما، على سبيل المثال، عبر الأماكن أو البيئات، أو عبر أنواع مختلفة من التجميعة. ومن شــأن مثل بحوث المقارنة هذه أن تُمكن مــن تحديد العموميات والخصوصيات، وكذلك الطــرق الاتفاقية والارتباطية المتنوعة التي تتسع من خلالها التجميعة ليتم تشكيلها.

وهناك طرق عديدة يمكن من خلالها تفعيل مثل هذه البحوث، وهي تشــمل تحليلات البيانات، كما تمت مناقشــته في الفصل الســادس، وذلك من أجــل تحديد تركيب تجميعة ما والارتباطات والعلاقات بين العناصر. وســيكون لمثل هذا النهج فوائد عديدة وذلك من خلال استخدام قوة التنقيب عن البيانات والتعلم الآلي من أجل تحديد الأنماط داخل وعبر مجاميع كبيرة للبيانات، وكذلك من أجل تأســيس قوة نســبية للروابط. وهذا من شأنه أن يوفر تفسيراً واسعاً ومقارناً للمجاميع وتشغيل هذه المجاميع. وعلى الرغم من ذلك، وفيما يتعلق بفهم الســياق، والاتفاقية والارتباطية المتعلقة بتجميعة ما -التداخل المعقد للكائن ومحيطه - فإنه من المحتمل، ربما بشكل ساخر، أن مجموعة من المنهجيات الخاصة بالبيانات الصغيرة قد توفر تبصراً إدراكياً تنويرياً دقيقاً. وتشــتمل هذه المنهجيات، على ســبيل المثال لا الحصر، على منهج علم الأصول (Genealogies)، ومنهج التفكيك (Deconstruction)، ومنهــج الأعراق والثقافــات (Ethnographies) ومنهج مشــاركة الملاحــظ (Observant Participation)، وهــي تُمثل مجموعة من المنهجيات التي أيدت اســتخدامها في الســابق





مــن أجل فهم البرامج (Kitchin and Dodge 2011)، والتخطيط (Dodge et al. 2009;) (Kitchin et al. 2012 a). وتسعى كل منهجية إلى أن تحدد وتفكك بشكل دقيق الظواهر الاجتماعيــة، وكما بينت العناصر الأحد عشر، فإن مجاميع البيانات تعتبر اجتماعية بشــكل كلي. ومن خلال اســتخدامها كمجموعة، فإن الإنتاج المادي والاســتطرادي المنتشر لمجاميع البيانــات يمكن تجزئته لأغــراض علمية، مما يوفر ما أطلق عليــه (2013) Wang مصطلح "البيانات الكثيفة"، أي الأفكار والقصص السياقية عن البيانات وإنتاج البيانات.

إن علم الأصول يستخدم كثيراً من أجل تتبع الانتشار العرضي لنظام الأفكار أو لمجموعة مــن الأفعــال عبر الأوقات والأزمــان، وذلك بدلاً مــن إنتاج تأريخ غائي ســليم (Crowley 2009). وهذا يوضح كيف أن المســتقبل يُبنى عــلى الماضي، ولكنه لا يتحدد بالضرورة من خلاله بطرق ســببية مســببيه بسيطة. وعلى هذا النحو، يتم توظيف علم الأصول من أجل الحل والفهم بشــكل تاريخي للتكرارات المتناقضة والمتغايرة أحيانــاً، والمعقدة، والمتعددة لأحد مجاميــع البيانات، أو أحد عناصرها وكذلك الفهم للانتشــار المتطور والقائم للأفكار، والقرارات، والمعوقات، والأفعال، واللاعبين الذين أســهموا في تطور هذه المجاميع، جنباً إلى جنب مع النهايات المســدودة والإخفاقات الواضحة (Foucault 1977). وعند فعل ذلك، فإن هذا يحدد نقاط التجمع حيث يحتشــد الناس أو الأفكار ســوياً ويتســببون في ظهور مجاميع جديــدة، وكذلك الطرق المعقدة والفوضوية التي تتطور من خلالها هذه المجاميع بعــد ذلــك. وفي حال ثورة البيانات، يمكن توظيف منهج علم الأصول من أجل تتبع تكوين وتطــور البيانات المفتوحة والبنى التحتية للبيانات والبيانات الكبيرة بشــكل أكثر عمومية، والتجسيد المحدد لها.

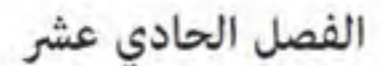
فعلى سبيل المثال، يمكن لشخص ما أن ينتج تأصيلاً مفصلاً لتطور حركة البيانات المفتوحة بتعبــيرات عامة، من خــلال تتبع ظهور اللاعبــين، والمنظمات، ونماذج التمويــل، والأفكار، والأحــداث بمرور الوقت وعبر الزمن، والترابطــات والتحالفات والخلافات بينهم، وتأثيراتهم على إعادة تشكيل مشهد البيانات من خلال إعادة تحديد السياسات التي تحيط بالحكومة التي أنتجت، واحتفظت، وكشفت مثل هذه البيانات من أجل استخدامات أعم وأشمل. أو يمكن للشــخص أن يشرع في تحديد أصل كيان فردي مثل مؤسســة المعرفة المفتوحة، وهي



مؤسســة غير ربحية تأسســت في العام 2004 في المملكة المتحدة، ونمت بسرعة لكي تصبح حركة دولية مؤثرة تطالب بالبيانات المفتوحة والحكومة المفتوحة. وهنا، يمكن أن يســعى الفرد تجاه تتبع نشــأة المنظمة، وتطور مهمتها وقيمها، وتطــور البنى التحتية، والحوكمة، والتمويل، والأنشطة، والبرامج، والإصدارات، وانتشار تأثيرها وتفاعلاتها مع الهيئات الأخرى وهكــذا، وذلك من أجل فهم الطريقة الاتفاقية والارتباطية لتوســعها وعملها على ســبيل المثال، وكذلك موقعها الحالي وسلطتها ضمن حركة البيانات المفتوحة الأعم.

إن إجراء تحليل علم الأصول عادةً ما يصاحبه الاستفادة من التفكيك، وهذه الطريقة من التحليل تجرد وتكشف المعاني داخل النصوص المفهومة بشكل واسع مثل النصوص المكتوبة، والصور، والخرائط، والأحاديث، ومزيج من هذه الأشــياء تشكل في العادة المحتوى للكتيبات والمواقــع الإلكترونية. فمن المفهوم أن النصوص هي الوســيط لكلٍ من الرســائل الصريحة والضمنية، ومن خلال التفسير التشريحي للنص، يمكن فهم والكشف عن معانٍ أعمق. ومن ناحية الجوهر، فإن التفكيك يبحث فيما يتجاوز الإشــارات الســطحية والقيم الاســمية لكي يواجه القراءات المسلم بها للنص وذلك من أجل الكشف عن المعاني المخفية بداخله، أو ربط المعاني المبسطة بشكل محكم (Burman and MacLure 2005; Harley 1989). فالتفكيك هو قراءة نقدية تســتخدم أســاليب مثل استبدال المعاني أو القوة المحتملة، أو تحديد نقاط التناقــض أو التباين لأجل حل النصيــة البينية والتضمين داخل الماديات والســياقات الأعم، والتفكير في وضعية المتحدث والجمهور المســتهدف له (Burman and MacLure 2005). وكــما أشرنــا إلى ذلك بالتفصيل في الفصول الثاني والثالث والســابع، فــإن البيانات المفتوحة والبنــى التحتية للبيانات والبيانــات الكبيرة عادة ما يواكبها جميعاً أنظمة خطابية تســعى إلى إقناع الناس، والشركات، والمؤسســات بمنطقها من أجل دعم وتبني والاستثمار في مبادئها ومنتجاتها. فالأنظمة الخطابية هذه ليست عالمية ولكنها معقدة وفوضوية، فهي مزيج من المصالح المتعاونة والمتنافسة. ومن خلال استخدام التفكيك، فإن الأنظمة الخطابية لتجميعة مــا أو عناصر معينة في هذه التجميعــة، مثل الفئات العديدة في حركة البيانات المفتوحة أو شركات معينــة تعمل على تعزيز حلول البيانات الكبيرة على القضايا المدنية، يمكن تجريدها وتفكيك رسالتها وتكوينها، مما يكشف عن المعاني، والأيدولوجيات (المعتقدات)، والسلطات الكامنة داخل تصميمها وعرضها.





وتوفـر دراسـات الأعراق والثقافات تحليلات شـاملة وهائلة لمجتمعات الممارسـة من خـلال التوثيق التفصيلي لتنظيمها، وعملها، وثقافتهـا، وقواها المحركة الداخلية وارتباطاتها الخارجيـة (Herbert 2000). وعـلى نحـو نموذجـي، يوظف المنهـج الإثنوغرافي (وصف الأعراق والثقافات) مزيجاً من مراقبة المشـاركين التي يتم القيام بها عبر فترة زمنية ممتدة، ومقابلات شـخصية متعمقة مع مجموعة كبيرة من المستفيدين الداخليين والخارجيين، مع تعزيز كل ما سـبق بأسـاليب أخرى مثل القراءات التفسيرية للوثائق والأدوات ذات الصلة مثل المواقـع الإلكترونية، وتبادلات البريد الإلكتروني، وأماكـن العمل... إلخ. ومن الناحية الجوهرية، يسـعى علم دراسـة الأعراق والثقافات إلى الفهم الدقيق للعالم الحياتي لمجتمع مـا - علاقاته الاجتماعية، وإيقاعاتـه المتواترة، ومعانيه الثقافية، وأنماط السـلطة، واتخاذ القرار فيه، وطرق كينونته،... إلخ - من أجل فهم كيف تشـكل هذا المجتمع وكيف ينمو بشـكل مسـتمر (Crang and Cook 2007). ويذهـب الباحث بعيداً متجـاوزاً تقديم وصف سـطحي لتوثيق الطرق المعقدة التي تعمل من خلالها المجتمعات من خلال كونها متضمنة داخل العلاقات والممارسات التي يتم دراستها. ومن شأن دراسة الأعراق والثقافات لمجاميـع البيانات أن تقدم بعد ذلك أوصافاً تفصيلية شـاملة عن كيف تم تكوينها وكيف تعمل على أرض الواقع. فعلى سـبيل المثال، فإن دراسة الأعراق والثقافات لمؤسسة المعرفة المفتوحة سوف تتكون من باحث تم تضمينه داخل تلك المنظمة، وهو يشارك في التخطيط للاجتماعات، والأنشـطة، والفعاليات، ويراقب تفاعلات العمال ويجري مقابلات شـخصية مع المستفيدين الأساسـيين للحصول على مدارك مباشرة حول عمليات اللاعبون الأساسيون في حركة البيانات المفتوحة، وكيف يتم تأسـيس وإدارة المكونات العديدة لمجاميع البيانات (بمعنى أنظمة الأفكار؛ والاقتصاد السـياسي؛ وأنماط الحاكميـة؛ والماديات؛ والبنى التحتية؛ إلخ) داخلها وعبرها. وعلى نحو مشـابه، يمكن للفرد أن يجري دراسـة الأعراق والثقافات داخل بنية تحتية معينة للبيانات، أو على سمسار للبيانات، أو داخل شركة للبيانات الكبيرة، أو داخل أنواع معينة من الفعاليات مثل سباقات البرمجة أو المعارض والمؤتمرات الصناعية، أو فرق البيانات الكبيرة داخل الحملات الانتخابية... إلخ.

ويمكن أن يتم دعم مثل هذه الدارسـات للأعراق والثقافات من خلال مشـاركة الملاحظ، وهذا نوع من الدراسـة الذاتية للأعراق والثقافـات حيث يُجري الباحث



دراسة مستديمة لارتباطاتهم وارتباطات الناس الآخرين بظاهرة أو ممارسة ما (Crang and Cook 2007; Morton 2005). وفيما يتعلق بثورة البيانات، من الممكن أن يتم مشاركة الملاحظين بحيث يكونون جزءاً من فريق يبني البنى التحتية للبيانات، أو يكونون أعضاءً في حركة البيانات المفتوحة، أو يعملون كعلماء للبيانات يوظفون البيانات الكبيرة من أجل معالجة قضية معينة... إلخ. فتوظيفٌ مثل هذا النهج يعتبر في الأصل تمرينًا للانعكاس الذاتي، وهو تدريب شخصي وغير موضوعي، حيث يسعى الباحث جاهداً وبدقة بالغة لدراسة ممارساته وممارسات الآخرين في مجال ما حيث يكون لاعباً أساسياً في هذا المجال، فالفائدة الأساسية لمثل هذا المنهج تتمثل في أن الباحث يدرك بشكل كامل المشهد المتنوع والمعقد من النواحي الاجتماعية، والسياسية، والاقتصادية والذي يعمل من خلاله هذا الباحث ويعتبر أحد المشاركين الفاعلين فيه ويتعرض للعمليات المتعددة الجارية. ويظل الانتقاد الأساسي الموجه لمثل هذا المنهج هو أنه من الممكن بدرجة كبيرة أن ينثني الباحث ويوضع في مكان ما بشكل شخصي، مما يفقده التمثيل العريض، وربما تتم الاستفادة من هذا المنهج بشكل أمثل إذا ما تم دمجه مع طرق أخرى توفر سياقاً أعم وأشمل.

إن هذه الطرق تمثل مجموعة جزئية من مجموعة من الطرق المحتملة لفهم مجاميع البيانات وعناصر هذه المجاميع، ومما لا شك فيه أن الكثير من هذه المناهج سوف يتم توظيفها من أجل إجراء بحوث تجريبية على طبيعة ثورة البيانات المنتشرة، إذ يوفر كل منهج رؤى مختلفة. وفي الواقع، فإن المنهج الذي يستخدم طرقاً مختلطة تمزج بين الرؤى السياقية المتعمقة للدراسات النوعية التي تستهدف البيانات الصغيرة وبين الدارسات الأعم التي توظف تحليلات البيانات ربما يكون سبيلاً مفيداً لاتباعه. وفي مثل هذه الدراسة، من الممكن استخدام تحليلات البيانات من أجل تقديم منظور شامل لمجاميع البيانات، أو الروابط أو التداخلات بين المجاميع، والتي يتم اتباعها بعد ذلك من خلال البحوث الأكثر تركيزاً باستخدام الطرق التي تم تفصيلها من أجل توفير رؤى متعمقة للعلاقات والعمليات السارية داخل وبين هذه العناصر. أو من الممكن أن يكون هذا المزيج مزيجاً كمياً صرفاً فيما يتعلق بالصيغة، وذلك على النحو الذي تمت الإشارة إليه بالتفصيل في المناقشات بشأن علم البيانات في الفصل الثامن، حيث تم تبني التحليل التفسيري لتجميعة البيانات بشكل

أكـبر من اختبار الفرضيات التقليدي من أجـل تحديد أكثر العناصر الفاعلة ظهوراً. ويمكن أن يعمل هذا المنهج أيضًا بشـكل عكسي، حيث يتم استخدام البحوث المتعمقة المركزة من أجل توجيه انتشـار تحليلات البيانات عبر عينة أكبر بكثير. وفي كلتا الحالتين، فإن المرحلتين الأولى والثانيـة تعمـلان بتناغم بعضهما ا مع بعض من أجل تعزيز المدارك والمعارف. ففي الحالة الأولى، يتم استخدام الصورة الأعم من أجل المساعدة في صياغة الفهم وتعميقه وقرنه بالسـياق، أما في الحالة الثانية، فإن الفهم المتعمق يوفر الأسـاس لمحاولة تأسـيس أو وضع تفسير أعم وأشمل.

أفكار ختامية:

لا تـزال ثورة البيانات في طورهـا التمهيدي، ولكنها تنتشر بسرعـة بالغة، فخلال عدد من السـنوات لا يتجـاوز أصابع اليد الواحـدة، كان للبيانات المفتوحـة والبيانات الكبيرة والبنـى التحتية للبيانات تأثيرات كبيرة على مشـهد البيانات، وتزامن هذا مع مناقشـات صوتية حماسية تعلن عن التأثيرات الإيجابية الخطيرة لها. إن سرعة التطورات التي تحدث تعني أن التحليل النقدي الموسـع، والمتعمق، والمدروس يجد صعوبة بالغة في مواكبة تلك التطورات. ومما لا شـك فيه أنه خلال السـنوات القليلة القادمة سـيكون هناك فيضان حقيقي من مثل هذه الدراسـات التي تسـعى إلى توثيق طبيعة وآثـار مجاميع البيانات الناشـئة. وكما ادعى البعض خلال هذا الفصل، فهناك حاجة ملحة لهذه الدراسـات. إن هـذا العمل سـوف يحدد من جانب الطرق المناسـبة لفهم ثورة البيانـات، من خلال كلٍ من التفكير المفاهيمي والدراسـات النقدية. وعلى الجانب الآخر، سيوفر بيئة لتفكير أكثر معيارية فيما يتعلق بكيفية تطور وتنظيم وإدارة مشهد البيانات. وفي الوقت الراهن، فإن الأخير يكون في الغالب تفاعلياً، ومفهومًا بشـكل ضعيف، وتسيطر عليه المصالح المكتسبة. لقد سـعى هذا الكتاب إلى تقديم تحليل أولي ومفاهيمي بشـكل متوازن، ونقدي وشامل والذي من الممكن أن يسـاعد على توجيه الدارسـات التي تتبعـه. ولوقت طويل للغاية، فإن البيانات وتكوين وعمل المجاميع التي تحيط بالبيانات قد تم الأخذ بها على أنها أمورٌ مسلم بها، في حين أن الاهتمام كان منصباً على المعلومات والمعارف المستقاة منها، فالوقت قد حان لتدارك هذا الإهمال.



المراجع:

- Adler, M.J. (1986) A Guidebook to Learning: For a Lifelong Pursuit of Wisdom. Macmillan, London.
- Agre, P. (1994) 'Surveillance and capture: two models of privacy', Information Society,10(2): 101–27.
- Alder, K. (2002) The Measure of All Things – The Seven-Year-Odyssey that Transformed the World. Abacus, London.
- Alonso, W. and Starr, P. (1987) The Politics of Numbers. Russell Sage Foundation, New York.
- Althusser, L. (1971) Lenin and Philosophy and Other Essays, translated by B. Brewster. New Left Books, London.
- Amin, A. and Thrift, N. (2002) Cities: Reimagining the Urban. Polity, London.
- Amin, A. and Thrift, N. (2013) Arts of the Political: New Openings for the Left. Duke University Press, Durham, NC.
- Amoore, L. (2006) 'Biometric borders: governing mobilities in the war on terror', Political Geography, 25: 336–51.
- Anderson, B. (2009) 'Affective atmospheres', Emotion, Space and Society, 2: 77–81.
- Anderson, C. (2008) 'The end of theory: the data deluge makes the scientific method obsolete', Wired, 23 June, http://www.wired.com/science/discoveries/magazine/1607-/pb_theory (last accessed 12 October 2012).
- Andrejevic, M. (2007) iSpy: Surveillance and Power in the Interactive Era. University of Kansas, Lawrence, KS.
- Andrejevic, M. (2013) Infoglut: How Too Much Information is Changing the Way We Think and Know. Routledge, New York.
- Anthony, S. (2013) 'DARPA shows off 1.8-gigapixel surveillance drone, can spot a ter¬rorist from 20,000 feet', ExtremeTech, 28 January 2013, http://www.extremetech.com/extreme/146909-darpa-shows-off-18--gigapixel-surveillance-drone-can-spot-



a-terrorist-from-20000-feet (last accessed 1 February 2013).

- Arthur, C. and Cross, M. (2006)' Give us back our crown jewels', The Guardian, 9 March. http://www.theguardian.com/technology/2006/mar/09/education.epublic (last accessed 21 October 2013).
- Asay, M. (2013) 'Big data and the landfills of the digital enterprise', Readwrite, 11 February. http://readwrite.com/201311/02//big-data-and-the-landfills-of-our-digital-lives (last accessed 19 February 2013).
- Bamford, J. (2009) The Shadow Factory: The Ultra-secret NSA from 911/ to the Eavesdropping on America. Anchor Books, New York.
- Barnes, T.J. and Hannah, M. (2001) 'The place of numbers: histories, geographies, and theories of quantification', Environment and Planning D: Society and Space, 19: 379–83.
- Barnett, C. (2013) 'Theory as political technology', Antipode (online first), http://radicalantipode.files.wordpress.com/201307//book-review_barnett-on-amin-and-thrift.pdf (last accessed 2 August 2013).
- Bates, J. (2012) '"This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data Initiative', The Journal of Community Informatics, 8(2) http://www.cijournal.net/index.php/ciej/article/view/845916/ (last accessed 6 February 2013).
- Bates, J. (2013) 'Opening up public data', SPERI Comment, 21 May. http://speri.dept.shef.ac.uk/201321/05//opening-public-data/ (last accessed 18 September 2013).
- Batty, M. (2007) Cities and Complexity: Understanding Cities with Cellular Automata, Agent Based Models, and Fractals. MIT Press, Cambridge, MA.
- Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G. and Portugali, Y. (2012) 'Smart cities of the future', European Physical Journal Special Topics, 214: 481–518.
- Baym, N.K. (2013) 'Data not seen: the uses and shortcomings of social media metrics',FirstMonday, 18(10), http://firstmonday.org/ojs/index.php/fm/article/view/48733752/(last accessed 3 January 2014).



- Beagrie, N., Lavoie, B. and Woollard, M. (2010) Keeping Research Data Safe 2, JISC, London and Bristol. http://www.beagrie.com/jisc.php.
- Becker, H. (1952) 'Science, culture, and society', Philosophy of Science, 19(4): 273–87.
- Beer, D. and Burrows, R. (2007) 'Sociology and, of and in Web 2.0: some initial considera¬tions', Sociological Research Online, 12(5), http://www.socresonline.org.uk/1217/5/.html.
- Beer, D. and Burrows, R. (2013) 'Popular culture, digital archives and the new social life of data', Theory, Culture and Society, 30(4): 47–71.
- Benkler, Y. (2006) The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven, CT.
- Berners-Lee, T. (2009) Linked Data, http://www.w3.org/DesignIssues/LinkedData.html (last accessed 30 January 2013).
- Berry, D. (2011) 'The computational turn: thinking about the digital humanities', Culture Machine, 12, http://www.culturemachine.net/index.php/cm/article/view/440470/ (last accessed 3 December 2012).
- Bertolucci, J. (2013) 'IBM, universities team up to build data scientists', Information Week, 15 January. http://www.informationweek.com/big-data/big-data-analytics/ibm-universities-team-up-to-build-data-scientists/ (last accessed 16 January 2014).
- Bettencourt, L.M.A., Lobo, J., Helbing, D., Kuhnert, C. and West, G.B. (2007) 'Growth, innovation, scaling, and the pace of life in cities', Proceedings of the National Academy of Sciences, 104(17): 7301–06.
- Bielsa, A. (2013) 'The Smart City Project in Santander', Sensors, 1 March,http://www.sensorsmag.com/wireless-applications/smart-city-project-santander-11152(last accessed 27 May 2013).
- Boellstorff, T. (2013) 'Making big data, in theory', First Monday, 18(10),http://firstmonday.org/ojs/index.php/fm/article/view/48693750/ (last accessed 3 January 2014).
- Bollier, D. (2010) The Promise and Peril of Big Data. The Aspen Institute.http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_ of_Big_Data.pdf (last accessed 1 October 2012).



- Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J. and Wilderman, C.C. (2009) Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education.http://www.informalscience.org/documents/PPSR report 20FINAL.pdf. A CAISE Inquiry Group Report, Center for Advancement of Informal Science Education (CAISE). Washington, DC (last accessed 17 January 2014).
- Borgman, C.L. (2007) Scholarship in the Digital Age. MIT Press, Cambridge, MA.
- Borgman, C.L. (2012) 'The conundrum of sharing research data', Journal of the American Society for Information Science and Technology, 63(6): 1059–78.
- Bowker, G. (2005) Memory Practices in the Sciences. MIT Press, Cambridge, MA.
- Bowker, G. and Star, L. (1999) Sorting Things Out: Classification and Its Consequences. MIT Press, Cambridge, MA.
- boyd, D. and Crawford, K. (2011) 'Six provocations for big data', SSRN,http://ssrn.com/abstract=1926431 (last accessed 22 February 2013)
- boyd, D. and Crawford, K. (2012) 'Critical questions for big data', Information, Communication and Society, 15(5): 662–79.
- Brooks, D. (2013a) 'The philosophy of data', New York Times, 4 February,http://www.nytimes.com/201305/02//opinion/brooks-the-philosophy-of-data.html (last accessed 5 February 2013).
- Brooks, D. (2013b) 'What data can't do', New York Times, 18 February,http://www.nytimes.com/201319/02//opinion/brooks-what-data-cant-do.html (last accessed 18 February 2013).
- Bruns, A. (2013) 'Faster than the speed of print: reconciling "big data" social media analysis and academic scholarship', First Monday, 18(10),http://firstmonday.org/ojs/index.php/fm/article/view/48793756/ (last accessed 3 January 2014).
- Bryant, R., Katz, R.H. and Lazowska, E.D. (2008) Big-Data Computing: Creating Revolutionary Breakthroughs in Commerce, Science and Society. December, Computing Community Consortium, http://www.cra.org/ccc/docs/init/Big_Data.pdf (last accessed 12 October 2012).



- Brynjolfsson, E., Hitt, L.M. and Kim, H.H. (2011) Strength in Numbers: How Does Data- Driven Decision-making Affect Firm Performance? SSRN,http://ssrn.com/abstract=1819486 (last accessed 22 February 2013).
- Burman, C. and MacLure, M. (2005) 'Deconstruction as a method of research', in B. Somekh and C. Lewin (eds), Research Methods in the Social Sciences. Sage, London, pp. 282–92.
- Callebaut, W. (2012) 'Scientific perspectivism: a philosopher of science's response to the challenge of big data biology', Studies in History and Philosophy of Biological and Biomedical Sciences, 43: 69–80.
- Calzolari, M.C. (2012) 'Analysis of Twitter followers of leading international companies', 8 June, http://www.camisanicalzolari.com/MCC-Twitter-ENG.pdf (last accessed 3 January 2014).
- Campbell, D.T. (1976) Assessing the Impact of Planned Social Change. The Public Affairs Center, Dartmouth College, New Hampshire.
- Carr, N.G. (2007) 'The ignorance of crowds', Strategy + Business Magazine, 47: 1–5.
- Carroll, R. (2013) 'Welcome to Utah, the NSA's desert home for eavesdropping on America', The Guardian, 14 June. http://www.guardian.co.uk/world/2013/jun/14/nsa-utah-data-facility (last accessed 8 July 2013).
- Case, D.O. (2002) Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Academic Press, Amsterdam.
- Castells, M. (1988) The Informational City: Information Technology, Economic Restructuring and the Urban-regional Process. Blackwell, Oxford.
- Castells, M. (1996) Rise of the Network Society. Blackwell, Oxford.
- Cavoukian, A. (2009) Privacy by Design: A Primer.http://www.privacybydesign.ca/content/uploads/201310//pbd-primer.pdf (last accessed 15 October 2013).
- CCSDS (2012) Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, Washington, DC.http://public.ccsds.org/publications/archive/650x0m2.pdf (last accessed 21 October 2013).



- Chignard, S. (2013) 'A brief history of open data', Paris Tech Review, 29 March,http://www.paristechreview.com/201329/03//brief-history-open-data/ (last accessed 18 September 2013).
- Chopra, S. and Meindl, P. (2012) Supply Chain Management: Strategy, Planning and Operation, 5th edition. Pearson, Harlow.
- Christensen, C.M. (1997) The Innovator's Dilemma. Harvard Business Review Press, Cambridge, MA.
- CIPPIC (2006) On the Data Trail: How Detailed Information about You Gets into the Hands of Organizations with whom You Have no Relationship. A Report on the Canadian Data Brokerage Industry. http://www.cippic.ca/uploads/May106-/DatabrokerReport.pdf, The Canadian Internet Policy and Public Interest Clinic, Ottawa (last accessed 17 January 2014).
- Clark, L. (2013) 'No questions asked: big data firm maps solutions without human input',Wired, 16 January, http://www.wired.co.uk/news/archive/201316/01-/ayasdi-big-data-launch (last accessed 28 January 2013).
- Clarke, R. (1988) 'Information technology and dataveillance', Communications of the ACM, 31(5): 498–512.
- Clarke, R. (1994a) 'The digital persona and its application to data surveillance', The Information Society, 10(2): 77–92.
- Clarke, R. (1994b) 'Human identification in information systems: management chal¬lenges and public policy issues', Information Technology and People, 7(4): 6–37.
- Claverie-Berge, I. (2012) 'Solutions big data IBM', IBM corporation. http://www-05.ibm.com/fr/events/netezzaDM_2012/Solutions_Big_Data.pdf (last accessed 9 July 2013).
- Clements, A. (2013) Democratic State Surveillance, Transparency and Trust. Cyberdialogue conference, University of Toronto, 17–18 March. http://www.cyberdialogue.ca/201303//democratic-state-surveillance-transparency-and-trust-by-andrew-clement/ (last accessed 8 July 2013).
- Clifford, S. (2012) 'Shopper alert: price may drop for you alone', New York



Times, 9 August, http://www.nytimes.com/201210/08//business/supermarkets-try-customizing-prices-for-shoppers.html (last accessed 8 August 2013).

- Codd, E.F. (1970) 'A relational model of data for large shared data banks', Communications of the ACM, 13(6): 377–87.
- Cohen, D. (2008) 'Contribution to: The Promise of Digital History' (round table discus¬sion), Journal of American History, 95(2): 452–91.
- Cohen, J. (2012) 'What is privacy for?' Social Sciences Research Network, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2175406 (last accessed 16 July 2013)
- Constine, J. (2012) 'How big is Facebook's data? 2.5 billion pieces of content and 500+ terabytes ingested every day', TechCrunch, 22 August, http://techcrunch.com/201222/08//how-big-is-facebooks-data-25--billion-pieces-of-content-and-500-terabytes-ingested-every-day/ (last accessed 28 January 2013).
- Coterill, C. (2011) 'Location privacy: who protects?', URISA, 23(2): 49–59.
- CRA (2003) Grand Research Challenges in Information Systems, The Computing Research Association. http://archive.cra.org/reports/gc.systems.pdf (last accessed 16 January 2014).
- Crampton, J., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M.W. and Zook, M. (2012) Beyond the Geotag? Deconstructing 'Big Data' and Leveraging the Potential of the Geoweb, http://www.uky.edu/~tmute2/geography_methods/readingPDFs/2012-Beyond-the-Geotag-2012.10.01.pdf (last accessed 21 February 2013).
- Crang, M. and Cook, I. (2007) Doing Ethnographies. Sage, London.
- Crawford, K. (2013) 'The hidden biases of big data', Harvard Business Review Blog, 1 April, http://blogs.hbr.org/201304//the-hidden-biases-in-big-data/ (last accessed 18 September 2013).
- Critical Art Ensemble (1995) 'The mythology of terrorism on the Net', http://www.critical-art.net/books/digital/tact2.pdf (last accessed 31 May 2013).
- Croll, A. (2012) 'Big data is our generation's civil rights issue, and we don't know



it', inBig Data Now: Current Perspectives from O'Reilly Media. O'Reilly Media, Sebastopol, CA, pp. 55–9.

- Crovitz, L.G. (2012) 'Obama's "Big Data" Victory', Wall Street Journal, 18 November,http://online.wsj.com/article/SB10001424127887323353204578126671124151266.html(last accessed 19 November 2012).
- Crowley, U. (2009) 'Genealogy method', in R. Kitchin and N. Thrift (eds), International Encyclopedia of Human Geography, vol. 4. Elsevier, Oxford, pp. 341–4.
- Crutcher, M. and Zook, M. (2009) 'Placemarks and waterlines: racialized cyberscapes in post-Katrina Google Earth', Geoforum, 40(4): 523–34.
- Cukier, K. (2010) 'Data, data everywhere', The Economist, 25 February, http://www.economist.com/node/15557443 (last accessed 12 November 2012).
- Culler, J. (2010) 'The closeness of close reading', ADE Bulletin, 149: 20–25.
- Curry, M.R. (1997) 'The digital individual and the private realm', Annals of the Association of American Geographers, 87: 681–99.
- Curry, M.R., Philips, D.J. and Regan, P.M. (2004) 'Emergency response systems and the creeping legibility of people and places', The Information Society, 20: 357–69.
- Cyberinfrastructure Council (2007) 'Cyberinfrastructure vision for 21st century discov-ery', http://www.nsf.gov/pubs/2007/nsf0728/index.jsp?org=EEC, National Science Foundation, Washington, DC (last accessed 17 January 2014).
- Danna, A. and Gandy, O.H. (2002) 'All that glitters is not gold: digging beneath the sur¬face of data mining', Journal of Business Ethics, 40: 373–86.
- Darnton, R. (2000) 'An early information society: news and the media in eighteenth-century Paris', American Historical Review, 105: 1–30.
- Dasish (2012) Roadmap for Preservation and Curation in the Social Sciences and Humanities, http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_ and_Curation_in_the_SSH.pdf (last accessed 15 October 2013).
- Data Seal of Approval (2010) Quality Guidelines for Digital Research Data,http://www.datasealofapproval.org/media/filer_public/201327/09//dsa-booklet_1_



june2010.pdf (last accessed 8 October 2013).

- DataRemixed (2013) Worldwide Open Data Sites, 8 August, http://dataremixed.com/201308//worldwide-open-data-sites (last accessed 4 November 2013).
- Davenport, T.H., Barth, P. and Bean, R. (2012) 'How "Big Data" Is Different', MIT Sloan Management Review, 30 July, http://sloanreview.mit.edu/the-magazine/2012-fall/54104/how-big-data-is-different/ (last accessed 28 January 2013).
- de Goes, J. (2013) '"Big data" is dead. What's next?', VB/Big Data, 23 February, http://venturebeat.com/201322/02//big-data-is-dead-whats-next/ (last accessed 26 February 2013).
- De Micheli, C. and Stroppa, A. (2013) 'Twitter and the underground market', 11th Nexa Lunch Seminar, 22 May, http://nexa.polito.it/nexacenterfiles/lunch-11-de_micheli-stroppa.pdf (last accessed 3 January 2013).
- de Vries, M., Kapff, L., Negreiro Achiaga, M., Wauters, P., Osimo, D., Foley, P., Szkuta, K., O'Connor, J. and Whitehouse, D. (2011) Pricing of Public Sector Information Study (POPSIS), http://epsiplatform.eu/sites/default/files/models.pdf (last accessed 11 August 2013).
- Deloite (2012) Open Data: Driving Ingenuity, Innovation and Growth. Deloitte, London. http://www.deloitte.com/assets/DcomUnitedKingdom/Local%20Assets/Documents/Market%20insights/Deloitte%20Analytics/uk-insights-deloitte-analytics-open-data-june-2012.pdf (last accessed 14 August 2013).
- Dembosky, A., Waters, R. and Steel, E. (2012) 'Questions raised over power of big data', Financial Times, Tech Hub, 11 December, http://www.ft.com/intl/cms/s/092/f76 77a-40dd-11e2-aafa-00144feabdc0.html (last accessed 11 December 2012).
- Dery, K., Hall, R. and Wailes, N. (2006) 'ERPs as technologies-in-practice: social con¬struction, materiality and the role of organisational factors', New Technology, Work and Employment, 21(3): 229–41.
- Desrosières, A. (1998) The Politics of Large Numbers: A History of Statistical Reasoning, trans¬lated by C. Naish. Harvard University Press, Cambridge, MA.



- Dicken, P. (2003) The Global Shift. Sage, London.
- Diebold, F. (2012) 'A personal perspective on the origin(s) and development of "big data": the phenomenon, the term, and the discipline',http://www.ssc.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf (last accessed 5 February 2013).
- Dietrich, D. (2012) 'Linked data', ePSIplatform Topic Report, no: 2012 11, http://epsiplatform.eu/content/topic-report-linked-data (last accessed 14 August 2013).
- Dodge, M. and Kitchin, R. (2004) 'Flying through code/space: the real virtuality of air travel', Environment and Planning A, 36: 195–211.
- Dodge, M. and Kitchin, R. (2005) 'Codes of life: identification codes and the machine-readable world', Environment and Planning D: Society and Space, 23(6): 851–81.
- Dodge, M. and Kitchin, R. (2007a) 'The automatic management of drivers and driving spaces', Geoforum, 38(2): 264–75.
- Dodge, M. and Kitchin, R. (2007b) '"Outlines of a world coming in existence": perva¬sive computing and the ethics of forgetting', Environment and Planning B, 34(3): 431–45.
- Dodge, M. and Kitchin, R. (2013) 'Mapping experience and knowledge: crowdsourced cartography', Environment and Planning A, 45(1): 19–36.
- Dodge, M., Perkins, C. and Kitchin, R. (2009) 'Mapping modes, methods and moments: a manifesto for map studies', in M. Dodge, R. Kitchin and C. Perkins (eds), Rethinking Maps. Routledge, London, pp. 220–43.
- Donovan, K. (2012) 'Seeing like a slum: towards open, deliberative development', Georgetown Journal of International Affairs, 13(1): 97–104.
- Dourish, P. (2001) Where the Action Is. MIT Press, Cambridge, MA.
- Dourish, P. and Bell, G. (2007) 'The infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space', Environment and Planning B, 34: 414–30.
- Driscoll, K. (2012) 'From punched cards to "big data": a social history of database pop¬ulism', communication +1, 1(4), http://scholarworks.umass.edu/cpo/vol1/



iss14/ (last accessed 22 February 2013).

- Dumbill, E. (2012) 'What is big data?' in Big Data Now: Current Perspectives from O'Reilly Media. O'Reilly Media, Sebastopol, CA. pp. 3–10.
- Dwoskin, E. (2013) 'Data broker discloses what it knows', Wall Street Journal, 4 September, http://blogs.wsj.com/digits/201304/09//data-broker-discloses-what-it-knows/(last accessed 11 October 2013).
- Dyche, J. (2012) 'Big data "eurekas!" don't just happen', Harvard Business Review Blog, 20 November, http://blogs.hbr.org/cs/201211//eureka_doesnt_just_happen.html(last accessed 23 November 2012).
- Eaves, D. (2013) 'The value of open data – don't measure growth, measure destruction', 25 April, http://eaves.ca/201325/04//the-value-of-open-data-dont-measure-growth-measure-destruction/ (last accessed 18 September 2013).
- Economist, The (2010) 'All too much: Monstrous amounts of data', 25 February, http://www.economist.com/node/15557421 (last accessed 12 November 2012).
- Economist, The (2013) 'Open data: a new goldmine', 18 May, http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innova¬tion-new-goldmine (last accessed 16 September 2013).
- Edwards, J. (2013) 'Facebook is about to launch a huge play in "big data" analytics',Business Insider, 10 May, http://www.businessinsider.com/facebook-is-about-to-launch-a-huge-play-in-big-data-analytics-20135- (last accessed 18 September 2013).
- Efrati, A., Thurm, S. and Searchy, D. (2011) 'Mobile-app makers face U.S. privacy inves¬tigation', Wall Street Journal, 5 April, http://online.wsj.com/article/SB10001424052748703806304576242923804770968.html(last accessed 17 July 2013).
- Elliott, C. (2004) 'Some rental cars are keeping tabs on the drivers', New York Times,13 January, http://www.nytimes.com/200413/01//business/13gps.html (last accessed 8 August 2013).
- Ellis, J. (2013) 'Hiding in public: how the National Archives wants to open up its data to Americans', Nieman Journalism Lab, 25 February,http://www.niemanlab.



org/201302//hiding-in-public-how-the-national-archives-wants-to-open-up-its-data-to-americans/ (last accessed 26 February 2013).

- Elwood, S. and Leszczynski, A. (2011) 'Privacy reconsidered: new representations, data practices, and the geoweb', Geoforum, 42: 6–15.
- European Commission (2012) Commission Proposes a Comprehensive Reform of the Data Protection Rules, 25 January, http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm (last accessed 6 August 2013).
- Farber, D. (2013) 'Counting the Internet of things in real time', CNet, 30 July, http://news.cnet.com/830176-57596162-3_11386-/counting-the-internet-of-things-in-real-time/ (last accessed 18 September 2013).
- Farber, M., Cameron, M., Ellis, C. and Sullivan, J. (2011) Massive Data Analytics and the Cloud: A Revolution in Intelligence Analysis. Booz Allen Hamilton. http://www.boozallen.com/media/file/MassiveData.pdf (last accessed 16 July 2013).
- Federal Trade Commission (2012) Protecting Consumer Privacy in an Era of Rapid Change, http://www.ftc.gov/os/2012120326/03/privacyreport.pdf (last accessed 14 October 2013).
- Ferro, E. and Osella, M. (2013) Eight Business Model Archetypes for PSI Re-Use.Open Data on the Web, Workshop, 23–24 April 2013, Google Campus, Shoreditch, London. http://www.w3.org/201304//odw/odw13_submission_27.pdf (last accessed 13 August 2013).
- Fitzgerald, A.M. (2010) State of Play: PSI Reuse in Australia. European Public Sector Information Platform, Topic Report No. 13, http://epsiplatform.eu/sites/default/files/ePSIplatform%20Topic%20Report%20No.%201320%Australia%20final.pdf (last accessed 14 August 2013).
- Flaounas, I., Ali, O., Lansdall-Welfare, T., De Bie, T., Mosdell, N., Lewis, J. and Cristianini, N. (2013) 'Research methods in the age of digital journalism', Digitial Journalism, 1(1): 102–16.
- Floridi, L. (2005) 'Is information meaningful data?', Philosophy and Phenomenological Research, 70(2): 351–70.



- Floridi, L. (2008) 'Data', in W.A. Darity (ed.), International Encyclopedia of the Social Sciences, 2nd edition. Detroit: Macmillan. Preprint online: http://www.philosophyofinformation.net/publications/pdf/data.pdf.
- Floridi, L. (2010) Information: A Very Short Guide. Oxford University Press, Oxford.
- Floridi, L. (2012) 'Big data and their epistemological challenge', Philosophy and Technology, 25(4): 435–7.
- Flowers, M. (2013) 'Beyond open data: the data-driven city', in B. Goldstein and L. Dyson (eds), Beyond Transparency: Open Data and the Future of Civic Innovation. Code for America Press, San Francisco, CA. pp. 196–209.
- Foucault, M. (1977) Discipline and Punish. Allen Lane, London.
- Foucault, M. (1978) The History of Sexuality, Volume One. Random House, New York.
- Foucault, M. (1981) Power/Knowledge: Selected Interviews and Other Writings 1972–1977. Harvester, London.
- Franks, B. (2012) Taming the Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics. Wiley, Hoboken, NJ.
- Fry, J., Lockyer, S., Oppenheim, C., Houghton, J.W. and Rasmussen, B. (2008) Identifying Benefits Arising from the Curation and Open Sharing of Research Data Produced by UK Higher Education and Research Institutes, JISC, London and Bristol.http://repository.jisc.ac.uk/279/ (last accessed 21 October 2013).
- Gannon, D. and Reid, D. (2009) 'Parallelism and the cloud', in T. Hey, S. Tansley and K. Tolle (eds), The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, WA, pp. 131–5.
- Gantz, J. and Reinsel, D. (2011) Extracting Value from Chaos. IDC, Framingham, MAhttp://idcdocserv.com/1142 (last accessed 1 October 2012).
- Gantz, J., Reinsel, D., Chute, C., Schlichting, W., McArthur, J., Minton, S., Xheneti, I., Toncheva, A. and Manfrediz, A. (2007) The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010. IDC, Framingham,



MA, http://www.emc.com/collateral/analyst-reports/expanding-digital-idc-white-paper.pdf (last accessed 26 October 2012).

- Garvey, E.G. (2013) '"facts and FACTS": abolitionists' database innovations', in L. Gitelman (ed), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 89–102.
- Gershenfeld, N., Krikorian, R. and Cohen, D. (2004) 'The internet of things', Scientific American, October, 291(4): 76–81.
- Gilder, G. (2000) Telecosm: How Infinite Bandwidth will Revolutionize Our World. Free Press, New York.
- Gilheany, S. (2000) 'Projecting the cost of magnetic disk storage over the next 10 years', http://www.teqnium.in/images/MagneticStorageOvertheNext10years.pdf (last accessed 16 January 2014).
- Gitelman, L. and Jackson, V. (2013) 'Introduction', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 1–14.
- Gleeson, J., Kitchin, R., Bartley, B. and Treacy, C. (2009) 'New ways of mapping social inclusion in Dublin City', http://www.nuim.ie/nirsa/mappinginclusion.pdf, Dublin City Partnership/Dublin City Development Board (last accessed 13 August 2013).
- Goddard, L. and Byrne, G. (2010) 'Linked data tools: semantic Web for the masses',First Monday, 15(11), http://firstmonday.org/ojs/index.php/fm/article/view/31202633/ (last accessed 14 August 2013).
- Goldberg, J. (2012) 'Big data ushers in era of security intelligence', Wired: Innovation Insights, 30 November, http://www.wired.com/insights/201211//big-data-ushers-in-era-of-security-intelligence/ (last accessed 3 December 2012).
- Golledge, R. and Stimson, R. (1997) Spatial Behaviour. Guilford, New York.
- González-Bailón, S. (2013) '"Big data" and the capillaries of human geography',Dialogues in Human Geography, 3(3): 292–6.
- González-Bailón, S., Wang, N., Rivero, A., Borge-Holtoefer, J. and Moreno, Y. (2012) Assessing the Bias in Communication Networks Sampled from Twitter. Working Paper, http://arxiv.org/abs/1212.1684 (last accessed 17 January 2014).



- Goodchild, M.F. (2007) 'Citizens as sensors: the world of volunteered geography',GeoJournal, 69: 211–21.
- Goodchild, M.F. (2009) 'Uncertainty', in R. Kitchin and N. Thrift (eds), International Encyclopedia of Human Geography, vol 12. Elsevier, Oxford, pp. 1–5.
- Gordon, G. (2013) 'Towards an overarching narrative on open government', Open Government Partnership UK, 12 February, http://www.opengovernment.org.uk/towards-an-overarching-narrative-on-open-government/ (last accessed 19 February 2013).
- Gorman, S. (2013) 'The danger of a big data episteme and the need to evolve geographic information systems', Dialogues in Human Geography, 3(3): 285–91.
- Goss, J, (1995) '"We know who you are and we know where you live": the instrumental rationality of geodemographics systems', Economic Geography, 71: 171–98.
- Gould, P. (1981) 'Letting the data speak for themselves', Annals of the Association of American Geographers, 71(2): 166–76.
- Graham, M. (2012) 'Big data and the end of theory?' The Guardian, 9 March, http://www.guardian.co.uk/news/datablog/2012/mar/09/big-data-theory (last accessed 12 November 2012).
- Graham, M., Stephens, M. and Hale, S. (2013) 'Mapping the geoweb: a geography of Twitter', Environment and Planning A, 45: 100–02.
- Graham, S. (2005) 'Software-sorted geographies', Progress in Human Geography,29(5): 562–80.
- Graham. S. and Marvin, S. (1996) Telecommunications and the City. Routledge, London.
- Graham, S. and Marvin, S. (2001) Splintering Urbanism. Routledge, London.
- Gralla, P., Sacco, A. and Faas, R. (2011) 'Smartphone apps: is your privacy protected?', ComputerWorld,http://www.computerworld.com/s/article/9218163/Smartphone_apps_Is_your_privacy_protected_(last accessed 17 July 2013).
- Grant, D., Harley, B. and Wright, C. (2006) 'Editorial introduction: the work and organ¬isational implications of Enterprise Resource Planning systems', New Technology, Work and Employment, 21(3): 196–8.



- Granville, V. (2013) 'The curse of big data', analyticbridge, 5 January, http://www.analyticbridge.com/profiles/blogs/the-curse-of-big-data (last accessed 28 January 2013).
- Greenfield, A. (2006) Everyware: The Dawning Age of Ubiquitous Computing. New Riders, Boston.
- Greenwald, G. and MacAskill, E. (2013) 'NSA Prism program taps in to user data of Apple, Google and others', The Guardian, 7 June, http://www.guardian.co.uk/world/2013/jun/06/us-tech-giants-nsa-data (last accessed 8 July 2013).
- Grimes, S. (2011) 'Unstructured data and the 80 percent rule', Clarabridge, http://clarabridge.com/default.aspx?tabid=137&ModuleID=635&ArticleID=551 (last accessed 4 March 2013).
- Grochowski, E. and Halem, R.D. (2003) 'Technological impact of magnetic hard disk drives on storage systems', IBM Systems Journal, 42: 338–46.
- Gurstein, M. (2011) 'Open data: empowering the empowered or effective data use for everyone', First Monday, 16(2), http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/33162764/ (last accessed 6 February 2013).
- Gurstein, M. (2013) 'Should "Open Government Data" be a product or a service (and why does it matter?)', Gurstein's Community Informatics, 3 February, http://gurstein.wordpress.com/201303/02//is-open-government-data-a-product-or-a-service-and-why-does-it-matter/ (last accessed 6 February 2013).
- Haklay, M. (2010) 'How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets', Environment and Planning B,37: 682–703.
- Hales, D. (2013) 'Lies, damned lies and big data', Aid on the Edge of Chaos, 1 February 2013, http://aidontheedge.info/201301/02//lies-damned-lies-and-big-data/ (last accessed 5 February 2013).
- Han, J., Kamber, M. and Pei, J. (2011) Data Mining: Concepts and Techniques, 3rd edition. Morgan Kaufmann, Waltham, MA.



- Hancke, G.P., de Carvalho e Silva, B. and Hancke Jr., G.P. (2012) 'The role of advanced sensing in smart cities', Sensors, 13: 393–425.
- Hannah, M. (1997) 'Space and the structuring of disciplinary power: an interpretative review', Geografiska Annaler Series B, Human Geography, 79: 171–80.
- Hannah, M. (2011) Dark Territory in the Information Age. Ashgate, Farnham.
- Haque, U. (2012) 'What is a city that it would be "smart"?', Volume #34: City in a Box http://volumeproject.org/blog/201221/12//volume-34-city-in-a-box/.
- Haraway, D. (1991) Simians, Cyborgs and Women: The Reinvention of Nature. Routledge, New York.
- Harcourt, B.E. (2006) Against Prediction: Profiling, Policing and Punishing in an Actuarial Age. Chicago University Press, Chicago.
- Harley, J.B. (1989) 'Deconstructing the map', Cartographica, 26: 1–20.
- Harris, D. (2012) 'Researchers mine 2.5m news articles to prove what we already know', Gigaom, 26 November 2012, http://gigaom.com/201226/11//researchers-mine-25-m-news-articles-to-prove-what-we-already-know/ (last accessed 28 January 2013).
- Hart, J.A., Reed, R.R. and Bar, F. (1992) 'The building of the Internet: implications for the future of broadband networks', Telecommunications Policy, 16: 666–89.
- Harvey, D. (1972) Social Justice and the City. Blackwell, Oxford.
- Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd edition. Springer, New York.
- Hausenblas, M. (2012) '5 * open data', http://5stardata.info/ (last accessed 30 January 2013).
- Helbig, N., Cresswell, A.M., Burke, G.B. and Luna-Reyes, L. (2012) The Dynamics of Opening Government Data: A White Paper. Centre for Technology in Government, State University of New York, Albany. http://www.ctg.albany.edu/publications/reports/opendata/opendata.pdf (last accessed 6 February 2013).
- Helland, P. (2011) 'If you have too much data, then "good enough" is good

enough', ACM Queue, https://queue.acm.org/detail.cfm?id=1988603 (last accessed 4 July 2013).

- Herbert, S. (2000) 'For ethnography', Progress in Human Geography, 24(4): 550–68.
- Hewitt, R. (2010) Map of a Nation: A Biography of the Ordnance Survey. Granta Books, London.
- Hey, T., Tansley, S. and Tolle, K. (2009) 'Jim Gray on eScience: A transformed scientific method', in T. Hey, S. Tansley and K. Tolle (eds), The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research, Redmond, WA, pp. xvii–xxxi.
- Hilbert, M. and López, P. (2011) 'The world's technological capacity to store, communi¬cate, and compute information', Science, 10(331): 703–05.
- Hill, D. (2013) 'On the smart city: or, a "manifesto" for smart citizens instead', City of Sound, 1 February, http://www.cityofsound.com/blog/201302//on-the-smart-city-a-call-for-smart-citizens-instead.html (last accessed 5 February 2013).
- Hollands, R.G. (2008) 'Will the real smart city please stand up?', City, 12(3): 303–20.
- Hon, W.K., Millard, C. and Walden, I. (2011) The Problem of 'Personal Data' in Cloud Computing – What Information is Regulated?, Queen Mary University of London, School of Law Legal Studies Research Paper No. 752011/, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1783577 (last accessed 16 July 2013).
- Houghton, J. (2011) Costs and Benefits of Data Provision. Report to the Australian National Data Service. Centre for Strategic Economic Studies, Victoria University. http://ands.org.au/resource/houghton-cost-benefit-study.pdf (last accessed 14 August 2013).
- Howe, J. (2008) Crowdsourcing. Random House, New York.
- Huijboom, N. and Van der Broek, T. (2011) 'Open data: an international comparison of strategies', European Journal of ePractice, 12 (March/April), http://www.epractice.eu/files/European%20Journal%20epractice%20Volume%2012_1.pdf(last accessed 15 August 2013).
- IBM (2012) What is Big Data?, http://www-01.ibm.com/software/data/bigdata/ (last accessed 12 November 2012).



- IBM (n.d.) Big Data at the Speed of Business http://www01.ibm.com/software/data/bigdata/industry.html (last accessed 9 July 2013).
- Innes, M. (2001) 'Control creep', Sociological Research Online, 6(3), http://www.socresonline.org.uk/63//innes.html (last accessed 14th January 2014).
- Issenberg, S. (2012) The Victory Lab: The Secret Science of Winning Campaigns. New York, Crown.
- InterPARES 2 (2013) Dictionary Definitions. InterPARES 2 Terminology Database: http://www.interpares.org/ip2/ip2_terminology_db.cfm (last accessed 21 October 2013).
- Jacobs, A. (2009) 'The pathologies of big data', ACM Queue, 1 July, http://queue.acm.org/detail.cfm?id=1563874 (last accessed 12 November 2012).
- Janssen, K. (2012) 'Open Government Data: right to information 2.0 or its rollback version?', ICRIWorking Paper, 82012/, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152566 (last accessed 14 August 2013).
- Jenkins, T. (2013) 'Don't count on big data for answers', Scotsman, 12 February, http://www.scotsman.com/the-scotsman/opinion/comment/tiffany-jenkins-don-t-count-on-big-data-for-answers-12785890- (last accessed 11 March 2013).
- Jensen, H.E. (1950) 'Editorial note', in H. Becker (1952), Through Values to Social Interpretation. Duke University Press, Durham, NC, pp. vii–xi.
- Jensen, D. (2000) Data Snooping, Dredging and Fishing: The Dark Side of Data Mining, http://www.martinsewell.com/datamining/Jens00.pdf (last accessed 20 February 2013).
- Johnson, S. (2003) 'Offloading your memories', The New York Times, 14 December, http:// www.nytimes.com/200314/12//magazine/14OFFLOADING.html (last accessed 16 January 2014).
- Johnson, J.A. (2013) From Open Data to Information Justice. Paper Presented at the Annual Conference of the Midwest Political Science Association, 13 April, Chicago, Illinois. http://papers.ssrn.com/abstract=2241092 (last accessed 16 August 2013).
- Kaye, K. (2012) 'Obama's approach to big data: do as I say, not as I do', AdAge, 16 November, http://adage.com/article/digital/obama-s-approach-big-data-i-i/238346/



(last accessed 19 November 2012).

- Keim, D., Kohlhammer, J., Ellis, G. and Mansmann, F. (2010) Mastering the Information Age – Solving Problems with Visual Analytics. Eurographics Association. http://www.vismaster.eu/book/ (last accessed 16 August 2013).
- Kelling, S., Hochachka, W., Fink, D., Riedewald, M., Caruana, R., Ballard, G. and Hooker, G. (2009) 'Data-intensive science: a new paradigm for biodiversity studies',BioScience, 59(7): 613–20.
- King, G. (2011) 'Ensuring the data-rich future of the social sciences', Science,331 (February): 719–21.
- Kinsley, S. (2011) 'Anticipating ubiquitous computing: logics to forecast technological futures', Geoforum, 42(2): 31–240.
- Kinsley, S. (2012) 'Futures in the making: practices to anticipate "ubiquitous computing"', Environment and Planning A, 44(7): 1554–69.
- Kitchin, R. (1996) 'Methodological convergence in cognitive mapping research: investi¬gating configurational knowledge', Journal of Environmental Psychology, 16(3): 163–85.
- Kitchin, R. (1998) Cyberspace: The World in the Wires. John Wiley and Sons, Chichester.
- Kitchin, R. (2006) 'Positivistic geography and spatial science', in S. Aitken and G. Valentine (eds), Approaches in Human Geography. Sage, London. pp. 20–9.
- Kitchin, R. (2013) 'Big data and human geography: opportunities, challenges and risks',Dialogues in Human Geography, 3(3): 262–7.
- Kitchin, R. (2014) 'The real-time city? Big data and smart urbanism', GeoJournal, 79 (1): 1–14.
- Kitchin, R. and Dodge, M. (2006) 'Software and the mundane management of air travel',First Monday, 11(9), http://firstmonday.org/issues/special11_9/kitchin/ (last accessed 6 March 2014).
- Kitchin, R. and Dodge, M. (2011) Code/Space: Software and Everyday Life. MIT Press, Cambridge, MA.



- Kitchin, R. and Fotheringham, A.S. (1997) 'Aggregation issues in cognitive mapping research', Professional Geographer, 49(3): 269–80.
- Kitchin, R. and Tate, N. (1999) Conducting Research in Human Geography: Theory, Methodology and Practice. Prentice Hall, Harlow.
- Kitchin, R., Bartley, B., Gleeson, J., Cowman, M., Fotheringham, S. and Lloyd, C. (2007) 'Joined-up thinking across the Irish border: making the data more compatible',Journal of Cross Border Studies, 2: 22–33.
- Kitchin, R., Gleeson, J. and Dodge, M. (2012a) 'Unfolding mapping practices: a new epistemology for cartography', Transactions of the Institute of British Geographers,38(3): 480–96.
- Kitchin, R., O'Callaghan, C., Boyle, M., Gleeson J. and Keaveney, K. (2012b) 'Placing neoliberalism: the rise and fall of Ireland's Celtic Tiger', Environment and Planning A,44: 1302–26.
- Koops, B.J. (2011) 'Forgetting footprints, shunning shadows: a critical analysis of the "right to be forgotten" in big data practice', SCRIPTed, 8(3): 229–56.
- Kourtit, K., Nijkamp, P. and Arribas-Bel, D. (2012) 'Smart cities perspective – a com¬parative European study by means of self-organizing maps', Innovation, 25(2): 229–46.
- Kuhn, T. (1962) The Structure of Scientific Revolutions. University of Chicago Press, Chicago.
- Lake, C. (2013) '24 beautifully-designed web dashboards that data geeks will love', 3 June, Econsultancy, http://econsultancy.com/ie/blog/6284424--beautifully-designed-web-dashboards-that-data-geeks-will-love (last accessed 12 November 2013).
- Laney, D. (2001) '3D data management: controlling data volume, velocity and variety',Meta Group, http://blogs.gartner.com/doug-laney/files/201201//ad9493-D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (last accessed 16 January 2013).
- Latour, B. (1989) Science in Action. Harvard University Press, Cambridge, MA.
- Lauriault, T.P. (2012) Data, Infrastructures and Geographical Imaginations: Mapping Data Access Discourses in Canada. Ph.D. Thesis, Carleton University, Ottawa.



• Lauriault, T.P., Craig, B.L., Taylor, D.R.F. and Pulsifier, P.L. (2007) 'Today's data are part of tomorrow's research: archival issues in the sciences', Archivaria, 64: 123–79.

• Lauriault, T.P., Hackett, Y. and Kennedy, E. (2013) Geospatial Data Preservation Primer. Ottawa, Hickling, Arthurs and Low.

• Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D. and Van Alstyne, M. (2009) 'Computational social science', Science, 323: 721–33.

• Lehning, M., Dawes, N., Bavay, M. Parlange, M., Nath, S. and Zhao. F. (2009) 'Instrumenting the Earth: next-generation sensor networks and environmental sci¬ence', in T. Hey, S. Tansley and K. Tolle (eds), The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research; Redmond, WA. pp. 45–51.

• Lehrer, J. (2010) 'A physicist solves the city', New York Times, 17 December, http://www.nytimes.com/201019/12//magazine/19Urban_West-t.html (last accessed 23 December 2013).

• Leonelli, S. (2012) 'Introduction: making sense of data-driven research in the biological and biomedical sciences', Studies in History and Philosophy of Biological and Biomedical Sciences, 43: 1–3.

• Leyshon, A. and Thrift, N. (1999) 'Lists come alive: electronic systems of knowledge and the rise of credit-scoring in retail banking', Economy and Society, 28: 434–66.

• Linehan, T.P. (1991) History and development of Irish population censuses', Journal of the Statistical and Social Inquiry Society of Ireland, XXVI(IV): 91–125.

• Lohr, S. (2012) 'Sure, big data is great. But so is intuition', New York Times, 29 December. http://www.nytimes.com/201230/12//technology/big-data-is-great-but-dont-for¬get-intuition.html (last accessed 3 January 2013).

• Lohr, S. (2013) 'SimCity, for real: measuring an untidy metropolis', New York Times, 23 February, http://www.nytimes.com/201324/02//technology/nyucenter-develops-a-science-of-cities.html (accessed 1 April 2013).

• Longo, J. (2011) '#OpenData: digital-era governance: thoroughbred or new public man¬agement Trojan horse?', PP+G Review, 2(2), http://ppgr.files.wordpress.



com/201105//longo-ostry.pdf (last accessed 16 September 2013).

- Loukides, M. (2010) 'What is data science?', O'Reilly Radar, 2 June, http://radar.oreilly.com/201006//what-is-data-science.html (last accessed 28 January 2013).
- Lucas, P., Ballay, J. and McManus, M. (2012) Trillions: Thriving in the Emerging Information Ecology. John Wiley, Hoboken, NJ.
- Lusk, G. (2013) 'Big. Bad. Big data', The Bubble Chamber, 13 February, http://thebubblechamber.org/201302//big-bad-big-data/ (last accessed 19 February 2013).
- Lyon, D. (2002) 'Everyday surveillance: personal data and social classifications', Information, Communication and Society, 5: 242–57.
- Lyon, D. (2003a) 'Surveillance as social sorting: computer codes and mobile bodies', in Lyon, D. (ed.), Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination. Routledge, London, pp. 13–30.
- Lyon, D. (2003b) 'Airports as data filters: converging surveillance systems after September 11th', Information, Communication and Ethics in Society, 1: 13–20.
- Lyon, D. (2007) Surveillance Studies: An Overview. Polity, Cambridge.
- Mann, S., Nolan, J. and Wellman, B. (2003) 'Sousveillance: inventing and using wearable computing devices for data collection in surveillance environments', Surveillance and Society, 1(3): 331–55.
- Manovich, L. (2011) 'Trending: the promises and the challenges of big social data', http://www.manovich.net/DOCS/Manovich_trending_paper.pdf (last accessed 9 November 2012).
- Manyika, J., Chiu, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Hung Byers, A. (2011) Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.
- Marche, S. (2012) 'Literature is not data: against digital humanities', Los Angeles Review of Books, 28 October, http://lareviewofbooks.org/article.php?id=1040&fulltext=1(last accessed 4 April 2013).
- Marchetti, N. (2012) 'In Portugal, a smart city from the ground up', Sustainable



Cities Collective, 7 June, http://sustainablecitiescollective.com/namarchetti/42438/portugal-smart-city-ground (last accessed 27 May 2013).

- Marz, N. and Warren, J. (2012) Big Data: Principles and Best Practices of Scalable Realtime Data Systems. MEAP edition. Manning, Shelter Island, New York.
- Mattern, S. (2013) 'Methodolatry and the art of measure: the new wave of urban data science', Design Observer: Places, 5 November, http://designobserver.com/places/feature/038174// (last accessed 15 November 2013).
- Mayer-Schonberger, V. and Cukier, K. (2013) Big Data: A Revolution that will Change How We Live, Work and Think. John Murray, London.
- McCandless, D. (2010) 'Data, information, knowledge, wisdom', Information is Beautiful,29 November, http://www.informationisbeautiful.net/2010/data-information-knowledge-wisdom/ (last accessed 24 July 2013).
- McClean, T. (2011) Not with a Bang but a Whimper: the Politics of Accountability and Open Data in the UK. Paper prepared for the American Political Science Association Annual Meeting. Seattle, Washington, 1–4 September 2011. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1899790 (last accessed 19 August 2013).
- McCreary, D. (2009) 'Entity extraction and the semantic web', Semantic Web, 12 January, http://semanticweb.com/entity-extraction-and-the-semantic-web_b10675 (last accessed 19 July 2013).
- McKeon, S.G. (2013) 'Hacking the hackathon', Shaunagm.net, 10 October, http://www.shaunagm.net/blog/201310//hacking-the-hackathon/ (last accessed 21 October 2013).
- McNay, L. (1994) Foucault: A Critical Introduction. Polity Press, Oxford.
- Miller, H.J. (2010) 'The data avalanche is here. Shouldn't we be digging?', Journal of Regional Science, 50(1): 181–201.
- Miller, H.J. and Han, J. (2009) 'Geographic data mining and knowledge discovery: an overview', in H.J. Miller and J. Han (eds), Geographic Data Mining and Knowledge Discovery. Taylor and Francis, London, pp. 3–32.
- Miller, P. (2001) 'Governing by numbers: why calculative practices matter', Social



Research, 68(2): 379–96.

- Miller, P. (2010) Linked Data and Government. ePSIplatform Topic Report no. 7, http://epsi-platform.eu/content/topic-report-no-7-linked-data-and-government-0 (last accessed 14 August 2013).
- Miller-Rushing, A., Primack, R. and Bonney, R. (2012) 'The history of public participa¬tion in ecological research', Frontier in Ecology and the Environment, 10(6): 285–90.
- Mims, C. (2013) 'Coming soon: the cybercrime of things', The Atlantic, 6 August, http://www.theatlantic.com/technology/archive/201308//coming-soon-the-cybercrime-of-things/278409/ (last accessed 7 August 2013).
- Minelli, M., Chambers, M. and Dhiraj, A. (2013) Big Data, Big Analytics. Wiley, Hoboken, NJ.
- Mitchell, M. (1996) An Introduction to Genetic Algorithms. MIT Press, Cambridge, MA.
- Mooney, P., Sun, H., Corcoran, P. and Yan, L. (2011) 'Citizen generated spatial data and information: risks and opportunities', Proceedings of the 2nd International Conference on Network Engineering and Computer Science (ICNECS 2011), Xi'an, Shaanxi, China,http://www.cs.nuim.ie/~pmooney/websitePapers/Mooney-et-all-Submitted-May2011.pdf(last accessed 28 January 2013).
- Moore, G.E. (1965) 'Cramming more components onto integrated circuits', Electronics, 38(8), http://download.intel.com/museum/Moores_Law/Articles-Press_Releases/Gordon_Moore_1965_Article.pdf (last accessed 20 May 2013).
- Moretti, F. (2005) Graphs, Maps, Trees: Abstract Models for a Literary History. Verso, London.
- Morozov, E. (2013) To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist. Allen Lane, New York.
- Morton, F. (2005) 'Performing ethnography: Irish traditional music sessions and new methodological spaces', Social and Cultural Geography, 6: 661–76.
- Najmi (2004) 'Generations of computer', Techi/Warehouse, 13 August, http://www.



techiwarehouse.com/engine/a046ee08/Generations-of-Computer (last accessed 20 May 2013).

- Narayanan, A. and Shmatikov, V. (2010) 'Privacy and security: myths and fallacies of "personally identifiable information"', Communications of the ACM, 53(6): 24–6.
- National Science Foundation (2012) Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA). Programme solicitation NSF 12499-http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf (last accessed 25 February 2013).
- NISO (National Information Standards Organization) (2004) Understanding Metadata. Bethesda, MD, http://www.niso.org/publications/press/UnderstandingMetadata.pdf (last accessed 28 June 2013).
- Nordhaus, W.D. (2002) The Progress of Computing, version 5.2.2, Yale University. http://www.econ.yale.edu/~nordhaus/homepage/prog_083001a.pdf (last accessed 17 January 2014).
- Northcutt, C. (2012) 'Open data creates value', NYI, 28 August, http://www.nyi.net/blog/201208//open-data-creates-value/ (last accessed 28 January 2013).
- OECD (2008) OECD Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information. OECD, http://www.oecd.org/internet/ieconomy/40826024.pdf (last accessed 14 August 2013).
- O'Carroll, A. and Webb, S. (2012) Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences. NUI Maynooth and Digital Repository of Ireland, Dublin.
- O'Carroll, A., Collins, S., Gallagher, D., Tang, J. and Webb, S. (2013) Caring for Digital Content, Mapping International Approaches. NUI Maynooth, Trinity College Dublin, Royal Irish Academy and Digital Repository of Ireland, Dublin.
- O'Neill, J. E. (1995) 'The role of ARPA in the development of the ARPANET, 1961–1972', IEEE Annals of the History of Computing, 17(4): 76–81.
- O'Reilly, T. (2005) What is Web 2.0: Design Patterns and Business Models for the



Next Generation of Software, 30 September, http://www.oreillynet.com/pub/a/oreilly/tim/news/200530/09//what-is-web-20.html (last accessed 17 January 2014).

- Ohlhorst, F.J. (2013) Big Data Analytics: Turning Big Data into Big Money. Wiley, New York.
- Open Data Center Alliance (2012) Big Data Consumer Guide. Open Data Center Alliance, http://www.opendatacenteralliance.org/docs/Big_Data_Consumer_Guide_Rev1.0.pdf (last accessed 11 February 2013).
- Open Knowledge Foundation (2012) Open Data Handbook Documentation, 1.0.0, http://opendatahandbook.org/ (last accessed 16 August 2013).
- Open Scientist (2013) 'The levels of citizen science involvement – part 1', 23 January, http://www.openscientist.org/201301//the-levels-of-citizen-science.html (last accessed 28 January 2013).
- Pentland, A. (2012) 'Reinventing society in the wake of big data', Edge, 30 August, http://www.edge.org/conversation/reinventing-society-in-the-wake-of-big-data (last accessed 28 January 2013).
- Pérez-Montoro, M.G. and Díaz Nafría, J.M. (2010) 'Data', in J.M. Díaz Nafría, M.G. Pérez-Montoro and F.S. Alemany (eds), Glossary of Concepts, Metaphors, Theories and Problems Concerning Information. León: Universidad de León. http://glossarium.bitrum.unileon.es/Home/dato/data (last accessed 13 January 2014).
- Piatetsky-Shapiro, G. (2012) 'From data mining to big data and beyond', Inside Analysis, 18 April, http://www.insideanalysis.com/201204//data-mining-and-beyond/(last accessed 20 February 2013).
- Picton, P.D. (2000) Neural Networks, 2nd edition. Palgrave Macmillan, Basingstoke.
- Plumridge, B. (2012) 'Big data in a big brave world', Wired, 21 November, http://www.wired.co.uk/news/archive/201221/11-/future-of-big-data? (last accessed 23 November 2012).
- Pollock, R. (2006) 'The value of the public domain', IPPR, http://www.ippr.org/pub¬lication/551526//the-value-of-the-public-domain (last accessed 13 August 2013).



- Pollock. R. (2009) The Economics of Public Information. Cambridge Working Papers in Economics 0920. http://www.econ.cam.ac.uk/research/repec/cam/pdf/cwpe0920.pdf(last accessed 13 August 2013).
- Poovey, M. (1998) A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society. University of Chicago Press, Chicago, IL.
- Porter, T.M. (1995) Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press, Princeton, NJ.
- Porway, J. (2013) 'You can't just hack your way to social change', Harvard Business Review Blog, 7 March, http://blogs.hbr.org/cs/201303//you_cant_just_hack_your_way_to.html (last accessed 9 March 2013).
- Prensky, M. (2009) 'H. sapiens digital: from digital immigrants and digital natives to digital wisdom', Innovate, 5(3), http://www.innovateonline.info/index.php?view=article&id=705 (last accessed 12 October 2012).
- Quintero, E. (2012) 'The data-driven education movement', Shanker Blog, 22 October,http://shankerblog.org/?p=7015 (last accessed 12 November 2012).
- Rajaraman, A., Leskovec, J. and Ullman, J.D. (2012) Mining of Massive Datasets. Available at http://infolab.stanford.edu/~ullman/mmds.html, also available from Cambridge University Press.
- Raley, R. (2013) 'Dataveillance and countervailance', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 121–46.
- Rambam, S. (2008) Privacy is Dead, Get Over It. Presentation at the Last Hope Conference, New York. http://www.youtube.com/watch?v=Vsxxsrn2Tfs (last accessed 15 October 2013).
- Ramirez, E. (2013) 'The privacy challenges of big data: a view from the lifeguard's chair', Technology Policy Institute Aspen Forum, 19 August,http//ftc.gov/speeches/ramirez/130819bigdataaspen.pdf (last accessed October 11 2013).
- Ramsay, S. (2003) 'Toward an algorithmic criticism', Literary and Linguistic Computing, 18(2): 167–74.



- Ramsay, S. (2010) Reading Machines: Towards an Algorithmic Criticism. University of Illinois Press, Champaign, IL.
- Rezendes, C.J. and Stephenson, W.D (2013) 'Cyber security in the internet of things',Harvard Business Review Blog, 21 June, http://blogs.hbr.org/cs/201306// cyber_security_in_the_internet.html (last accessed 7 August 2013).
- Rial, N. (2013) 'The power of big data in Europe', New Europe, 24 May, http://www.neurope.eu/article/power-big-data-europe (last accessed 27 May 2013).
- Ribes, D. and Jackson, S.J. (2013) 'Data bite man: the work of sustaining long-term study', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 147–66.
- Risen, J. and Lichtblau, E. (2013) 'How the US uses technology to mine more data more quickly', New York Times, 8 June,http://www.nytimes.com/201309/06//us/revelations-give-look-at-spy-agencys-wider-reach.html?hp&_r=3& (last accessed 8 July 2013).
- Ritzer, G. and Jurgenson, N. (2010) 'Production, consumption, prosumption: the nature of capitalism in the age of the digital "prosumer"', Journal of Consumer Culture, 10(1): 13–36.
- RLG and OCLC (2002) Trusted Digital Repositories: Attributes and Responsibilities. http://www.oclc.org/research/activities/trustedrep.html (last accessed 18 February 2013).
- Robinson, S. (2003) Simulation: The Practice of Model Development and Use. John Wiley & Sons, Chichester.
- Rogers, S. (2013) 'Twitter's languages of New York mapped', Guardian, 21 February, http://www.guardian.co.uk/news/datablog/interactive/2013/feb/21/twitter-lan¬guages-new-york-mapped (last accessed 3 April 2013).
- Rooney, B. (2012) 'Big data's big problem: little talent', Wall Street Journal: Tech Europe, 26 April, http://blogs.wsj.com/tech-europe/201226/04//big-datas-big-problem-little-talent/ (last accessed 12 November 2012).
- Rose, A. (2013) 'The internet of things has arrived – and so have massive security issues', Wired, 11 January, http://www.wired.com/opinion/201301//securing-the-internet-of-things/ (last accessed 7 August 2013).



- Rose, N. (1996) Inventing Our Selves: Psychology, Power and Personhood. Cambridge University Press, Cambridge.
- Rosenberg, D. (2013) 'Data before the fact', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 15–40.
- Rubenking, N.J. (2013) 'Privacy is dead. The NSA killed it. Now what?', PC Mag, http://www.pcmag.com/article20,2817,2424193,00/.asp (last accessed 15 October 2013).
- Rubinstein, I.S. (2013) 'Big data: the end of privacy or a new beginning?', International Data Privacy Law, online first, http://idpl.oxfordjournals.org/content/early/201324/01//idpl. ips036.short (last accessed 15 July 2013).
- Ruppert, E. (2012) 'The governmental topologies of database devices', Theory, Culture Society, 29: 116–36.
- Ruppert, E. (2013) 'Rethinking empirical social sciences', Dialogues in Human Geography, 3(3): 268–73.
- Salmon, F. (2014) 'Why the Nate Silvers of the world don't know everything', Wired, 7 January, http://www.wired.com/business/201401//quants-dont-know-everything/ (last accessed 8 January 2014).
- Salus, P. (1995) Casting the Net: From Arpanet to Internet and Beyond. Addison Wesley, Reading, MA.
- Sawyer, S. (2008) 'Data wealth, data poverty, science and cyberinfrastructure' ,Prometheus: Critical Studies in Innovation, 26(4): 355–71.
- Schnapp, J. and Presner, P. (2009) Digital Humanities Manifesto 2.0. http://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf (last accessed 13 March 2013).
- Scherer, M. (2012) 'Inside the secret world of the data crunchers who helped Obama win', Time: Swampland, 7 November, http://swampland.time.com/201207/11//inside-the-secret-world-of-quants-and-data-crunchers-who-helped-obama-win/ (last accessed 19 November 2012).
- Science Coalition (2013) Sparking Economic Growth: Companies Created from Federally Funded University Research, Fueling American Innovation and Economic



Growth, October 2013, http://www.sciencecoalition.org/reports/Sparking%20 Economic%20Growth%20FINAL%201013-21-.pdf (last accessed 12 January 2014).

- Schwartz, P.M. and Solove, D. (2011) 'The PII problem: privacy and a new concept of personally identifiable information', New York University Law Review, 86: 1814–94.
- Selisker, S. (2012) 'The digital inhumanities?', Los Angeles Review of Books, 5 November, http://lareviewofbooks.org/article.php?type=&id=1146 (last accessed 4 April 2013).
- Seni, G. and Elder, J. (2010) Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions. Morgan & Claypool, San Rafael, CA.
- Shah, N. (2013) 'Big data, people's lives, and the importance of openness' , DMLcentral,24 June, http://dmlcentral.net/blog/nishant-shah/big-data-peoples-lives-and-importance-openness (last accessed 25 July 2013).
- Shah, S., Horne, A. and Capellá, J. (2012) 'Good data won't guarantee good decisions',Harvard Business Review, 90(4): 23–5.
- Shneiderman, B. (1996) 'The eyes have it: a task by data type taxonomy for information visualizations', Proceedings IEEE Visual Languages, 96: 336–43.
- Short, J.E., Bohn, R.E. and Chaitanya, C. (2011) How Much Information? 2010: Report on Enterprise Server Information. Global Information Industry Center, UC San Diego.http://hmi.ucsd.edu/pdf/HMI_2010_EnterpriseReport_Jan_2011.pdf (last accessed 1 October 2012).
- Sicular, S. (2013) 'Big data is falling into the trough of disillusionment', Gartner, 22 January, http://blogs.gartner.com/svetlana-sicular/big-data-is-falling-into-the-trough-of-disillusionment/ (last accessed 26 February 2013).
- Siegel, E. (2013) Predictive Analytics. Wiley, Hoboken, NJ.
- Sigala, M. (2005) 'Integrating customer relationship management in hotel operations: managerial and operations implications', International Journal of Hospitality Management, 24(3): 391–413.
- Silver, N. (2012) The Signal and the Noise: The Art and Science of Prediction. Penguin, London.



- Singer, N. (2012a) 'You for sale: mapping, and sharing, the consumer genome', New York Times, 17 June, http://www.nytimes.com/201217/06//technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html (last accessed 11 October 2013).
- Singer, N. (2012b) 'F.T.C. opens an inquiry into data brokers', New York Times, 18 December, http://www.nytimes.com/201219/12//technology/ftc-opens-an-inquiry-into-data-brokers.html (last accessed 11 October 2013).
- Singer, N. (2012c) 'Mission control, built for cities: I.B.M. takes 'Smarter Cities' concept to Rio de Janeiro', New York Times, 3 March, http://www.nytimes.com/201204/03//busi¬ness/ibm-takes-smarter-cities-concept-to-rio-de-janeiro.html (last accessed 9 May 2013).
- Singh, R. (2012) 'Crowdsourced geospatial data', GIM International, 26(9): 26–31.
- Slee, T. (2012) 'Seeing like a geek', Crooked Timber, 25 June, http://crookedtimber.org/201225/06//seeing-like-a-geek/ (last accessed 18 September 2013).
- Smolan, R. and Erwitt, J. (2012) The Human Face of Big Data. Sterling, New York.
- Solove, D.J. (2006) 'A taxonomy of privacy', University of Pennsylvania Law Review,154(3): 477–560.
- Solove, D.J. (2007) '"I've got nothing to hide" and other misunderstandings of privacy',Social Sciences Research Network, http://ssrn.com/abstract=998565 (last accessed 16 July 2013).
- Solove, D. (2013) 'Privacy management and the consent dilemma', Harvard Law Review, 126: 1880–903.
- Stalder, F. (2002) 'Privacy is not the antidote to surveillance', Surveillance and Society,1(1): 120–4.
- Star, S.L. and Ruhleder, K. (1996) 'Steps Toward an Ecology of Infrastructure: design and access for large information spaces', Information Systems Research, 7(1): 111–34.
- Star, S.L. and Lampland, M. (2009) 'Reckoning with standards', in M. Lampland and S.L. Star (eds) Standards and Their Stories: How Quantifying, Classifying and Formalizing Practices Shape Everyday Life. Cornell University Press, Ithaca, NY, pp. 3–24.



- Starr, P. (1987) 'The sociology of official statistics', in W. Alonso and P. Starr (eds), The Politics of Numbers. Russell Sage, New York, pp. 7–58.
- Steadman, I. (2013) 'Big data and the death of the theorist', Wired, 25 January, http://www.wired.co.uk/news/archive/201325/01-/big-data-end-of-theory (last accessed 30 January 2013).
- Strasser, B.J. (2012) 'Data-driven sciences: from wonder cabinets to electronic databases', Studies in History and Philosophy of Biological and Biomedical Sciences,43: 85–7.
- Strasser, C. (2013) 'Closed data … excuses, excuses', Data Pub: California Digital Library, 24 April, http://datapub.cdlib.org/201324/04//closed-data-excuses-excuses (last accessed 18 September 2013).
- Strohm, C. and Homan, T. (2013) 'NSA spying row in Congress ushers in debate over big data', Bloomberg, 25 July, http://www.bloomberg.com/news/201325-07-/nsa-spying-row-in-congress-ushers-in-debate-over-big-data.html (last accessed 25 July 2013).
- Strom, D. (2012) 'Big data makes things better', Slashdot, 3 August, http://slashdot.org/topic/bi/big-data-makes-things-better/ (last accessed 24 October 2013).
- Sword, K. (2008) 'Contribution to: The Promise of Digital History' (roundtable discus¬sion), Journal of American History, 95(2): 452–91.
- Taleb. N. (2012) Antifragile: Things That Gain From Disorder. Random House, New York.
- Taleb, N. (2013) 'Beware the big errors of "big data"', Wired, 2 February, http://www.wired.com/opinion/201302//big-data-means-big-errors-people/ (last accessed 9 February 2013).
- TechAmerica Foundation (2012) Demystifying Big Data: A Practical Guide to Transforming the Business of Government. http://www.techamericafoundation.org/official-report-of-the-techamerica-foundations-big-data-commission (last accessed 25 February 2013).
- Tene, O. and Polonetsky, J. (2012) 'Big data for all: privacy and user control in the age

of analytics', Social Sciences Research Network, http://ssrn.com/abstract=2149364 (last accessed 15 July 2013).

- Thomas, J.J. and Cook, K.A. (2006) 'A visual analytics agenda', IEEE Computer Graphics & Applications, 26: 10–13.
- Thrift, N. (2004) 'Driving in the city', Theory, Culture & Society, 21(459–41 :(5/.
- Townsend, A. (2013) Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia. W.W. Norton & Co, New York.
- Trumpener, K. (2009) 'Critical response I. Paratext and genre system: a response to Franco Moretti', Critical Inquiry, 36(1): 159–71.
- Vanacek, J. (2012) 'How cloud and big data are impacting the human genome – touch¬ing 7 billion lives', Forbes, 16 April, http://www.forbes.com/sites/sap/201216/04// how-cloud-and-big-data-are-impacting-the-human-genome-touching-7-billion-lives (last accessed 27 February 2013).
- Varian, H. (1996) 'Differential pricing and efficiency', First Monday, 1(2–5),http://firstmonday.org/ojs/index.php/fm/article/view/473394/ (last accessed 8 August 2013).
- Verwayen, H., Arnoldus, M. and Kaufman, P.B. (2011) The Problem of the Yellow Milkmaid: A Business Mode Perspective on Open Metadata. Europeana, http://pro.europeana.eu/documents/8585662/cbf1f78-e0364088--af2594684-ff90dc5(last accessed 12 September 2013).
- Vis, F. (2013) 'A critical reflection on big data: considering APIs, researchers and tools as data makers', First Monday, 18(10), http://firstmonday.org/ojs/index.php/fm/article/view/48783755/ (last accessed 3 January 2013).
- von Baeyer, H.C. (2003) Information: The New Language of Science. Harvard University Press, Cambridge, MA.
- Wang, T. (2013) 'Big data needs thick data', Ethnography Matters, 13 May, http://ethnographymatters.net/201313/05//big-data-needs-thick-data/ (last accessed 3 January 2014).



- Weber, R.H. (2010) 'Internet of things – new security and privacy challenges', Computer Law & Security Review, 26: 23–30.
- Webster, T. (2011) 'Social media data dredging', BrandSavant, 22 March, http://brandsavant.com/social-media-data-dredging/ (last accessed 20 February 2013).
- Weinberger, D. (2011) Too Big to Know. Basic Books, New York.
- White House (2009) Open Government Directive. Executive Office of the President. http://www.whitehouse.gov/sites/default/files/microsites/ogi-directive.pdf (last accessed 19 September 2013).
- Williford, C. and Henry, C. (2012) One Culture: Computationally Intensive Research in the Humanities and Social Sciences. Council on Library and Information Resources, Washington DC. http://www.clir.org/pubs/reports/pub151 (last accessed 6 January 2014).
- Wilson, M. (2011) 'Data matter(s): legitimacy, coding, and qualifications-of-life',Environment and Planning D: Society and Space, 29: 857–72.
- Wired (2008) 'The petabyte age: because more isn't just more – more is different', 23 June, http://www.wired.com/science/discoveries/magazine/1607-/pb_intro (last accessed 9 November 2012).
- Wittgenstein, L. (1921[1974]) Tractatus Logico-Philosophicus, translated by D.F. Pears and B.F. McGuinness. Routledge, London.
- Wu, M. (2012) 'The big data fallacy and why we need to collect even bigger data', Tech Crunch, 25 November, http://techcrunch.com/201225/11//the-big-data-fallacy-data-%E289%%A0-information-%E289%%A0-insights/ (last accessed 28 January 2013).
- Wu, T. (2011) The Master Switch: The Rise and Fall of Information Empires. Vintage, New York.
- Wyly, E. (in press) 'Automated (post)positivism', Urban Geography.
- Yiu, C. (2012) A Right to Data: Fulfilling the Promise of Open Public Data in the UK. Policy Exchange Research Note, http://www.policyexchange.org.uk/



publications/category/item/a-right-to-data-fulfilling-the-promise-of-open-public-data-in-the-uk (last accessed 14 August 2013).

- Zelany, M. (1987) 'Management support systems: towards integrated knowledge manage¬ment', Human Systems Management, 7: 59–70.
- Zhang, J.., Hsu, M. and Li, L. (2001) 'Image mining: issues, frameworks and techniques', CiteSeerx, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.4.8726 (last accessed 17 January 2014).
- Zikopoulos, P.C., Eaton, C., deRoos, D., Deutsch, T. and Lapis, G. (2012)Understanding Big Data. McGraw Hill, New York





المترجم في سطور

د. محمد بن أحمد علي غروي

المؤهل العلمي:

- دكتوراه علوم المعلومات؛ جامعة نيويورك الحكومية، الولايات المتحدة الأمريكية، ٢٠١٢م.

الوظيفة الحالية:

أستاذ تقنية المعلومات المساعد في قطاع تقنية المعلومات بمعهد الإدارة العامة بالرياض.

إنجاز العديد من الدراسات العلمية منها:

1. Gharawi, M., & Khoja, M. (2015). Assessing Basic Computer Applications' Skills of College-Level Students in Saudi Arabia. International Journal of Social, Behavioral, Educational, Economic and Management Engineering Vol: 9, No: 4.
2. Gharawi, M., Estevez, E., Janowski, T (2014). Identifying government chief information officer education and training needs: the case of Saudi Arabia. DG.O 2014: 280- 289
3. Deosthali, K., Gharawi, M. & Potnis, D. (2014). Assessing Employee Satisfaction of the Electronic Training Program of the Gulf Cooperation Council's Countries. In Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (pp. 488- 497).
4. Gharawi, M., & Dawes, S. (2013). Exploring the Influence of Contextual Distances on Transnational Public Sector Knowledge Networks: A Comparative Study of AIRNow-I Shanghai and the Hajj-MDSS Initiatives. Paper presented at the 46th Hawaii International Conference on System Sciences (HICSS-46), Maui, Hawaii.
5. Gharawi, M., & Pardo, T. (2012). Factors Influencing Exchange Activities in Transnational Knowledge Networks: The Case of the Hajj-MDSS. ACM International Conference Proceeding Series, 322, 113- 122.

6. Djoko, S., Jong, J., Nam, T. & Gharawi, M. (2012). Building the Academic Community of E-Government Research on Cross-Boundary Information Integration and Sharing. Paper presented at the 45th Hawaii International Conference on System Sciences (HICSS-45), Maui, Hawaii.

7. Dawes, S., Gharawi, M., & Burke, B. (2012). Transnational Public Sector Knowledge Networks: Knowledge and Information Sharing in a Multidimensional Context. Government Information Quarterly, 29(1), S112–S120.

8. Dawes, S., Gharawi, M., & Burke, B. (2011). Knowledge and Information Sharing in Transnational Knowledge Networks: A Contextual Perspective. Paper presented at the 44th Hawaii International Conference on System Sciences (HICSS-44), Kauai, Hawaii.

9. Gharawi, M., & Dawes, S. (2010). Conceptualizing Knowledge and Information Sharing in Transnational Knowledge Networks. Paper presented at the 4th International Conference on Theory and Practice of Electronic Governance, Beijing, China.

10. Rorissa, Demissie D., A., Gharawi, M. (2010). A Descriptive Analysis of Contents of Asian e-Government. In E-Government Website Development: Future Trends and Strategic Models. Ed Downey, SUNY College at Brockport: USA.

11. Rorissa, A., Gharawi, M., Demissie D. (2010). A Tale of Two Continents: Contents of African and Asian E-government Websites. Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS-43), Koloa, Kauai, Hawai.

12. Gharawi, M., Pardo, T., & Guerrero, S. (2009). Issues and Strategies for Conducting Cross-National E- government Comparative Research. ACM International Conference Proceeding Series, 322, 163- 170.

13. Karshmer, A., and Gharawi, M. (2001). Computer Speech for Teaching Mathematics to the Blind, The International Conference for Education and Training, Prague, Czech Republic.

14. Montilva, J., Hamzan, K., and Gharawi, M. (2000).The Watch Model for Developing Business Software in Small and Midsize Organizations, Proceedings of the SCI, Volume 1, Orlando, Florida.

أبرز الأنشطة العملية:

١. عضو اللجنة الرئيسية لمبادرة البرنامج الوطني للتدريب عن بُعد، ٢٠١٦م - الآن.

٢. عضو الفريق الوطني لبناء إطار حوكمة التعليم والتدريب الإلكتروني بالمملكة العربية السعودية، ٢٠١٦م.

٣. عضو الفريق الدولي لدراسة احتياج المهارات الوظيفية، صندوق الموارد البشرية بالتعاون مع جامعة هارفرد، ٢٠١٥-٢٠١٧م.

٤. عضو الفريق الوطني لقياس التحول للتعاملات الحكومية الإلكترونية (قياس) ٢٠١٥-الآن.

٥. عضو الفريق الدولي لجائزة الإنجاز في التعاملات الحكومية الإلكترونية (إنجاز) ٢٠١٤م.

٦. عضو الجمعية الدولية للحكومة الرقمية (DG.O) ٢٠١١ - الآن.

٧. مدير مشروع إثراء وتنويع مصادر المعرفة العربية ٢٠١٦-الآن؛ معهد الإدارة العامة، الرياض.

٨. مشرف على برنامج التدريب الإلكتروني ٢٠١٢م -٢٠١٤م؛ معهد الإدارة العامة، الرياض.

٩. باحث ومدير برامج ٢٠٠٩م -٢٠١١م؛ مركز أبحاث التقنية الحكومية، نيويورك، الولايات المتحدة الأمريكية.

١٠. تنفيذ عدد من الاستشارات للجهات الحكومية بالمملكة العربية السعودية ٢٠٠١م -٢٠١٧م.

١١. تصميم وتطوير العديد من البرامج التدريبية في مجال تقنية المعلومات ٢٠٠١م-٢٠١٧م.

١٢. المشاركة في العديد من المحافل والمؤتمرات المحلية والإقليمية والدولية.



مراجع الترجمة في سطور

د. ماجد بن عبدالعزيز البريثن

المؤهل العلمي:

- الدكتوراة في مجال أنظمة المعلومات والتخطيط الإستراتيجي للمنشآت، جامعة سالفورد بريطانيا (٢٠١١).

الوظيفة الحالية:

- أستاذ مساعد بقسم نظم المعلومات - كلية علوم الحاسب والمعلومات - جامعة الإمام محمد بن سعود الإسلامية.

الخبرات العلمية والعملية:

- عمل عضو هيئة تدريس بقسم تقنية المعلومات والاتصالات بكلية الملك فهد الأمنية.
- عمل محاضراً غير متفرغ بمركز خدمة المجتمع بجامعة الإمام محمد بن سعود.
- الإشراف ومناقشة عدد من رسائل الماجستير.
- حضور ونشر عدد من المشاركات العلمية بعدد من المؤتمرات والمجلات العالمية.
- عمل مستشاراً بهيئة الهلال الأحمر السعودي، ٢٠١٧ - حتى الآن.
- عضو فرق عمل ضمان الجودة لشركة QSR-International فيما يخص ضمان جودة واختبار برمجياتها (NVivo) تطبيق إدارة وتحليل البيانات الوصفية.
- العمل مستشار لأمانة مدينة الرياض في تطوير مشروع مركز الاتصال ٩٤٠ .
- العمل مع إدارة تقنية المعلومات بالمؤسسة العامة للتقاعد على تحليل بعض المشاكل القائمة.
- العمل على تطوير أنظمة حاسوبية لشركة القصبي.
- العمل والإشراف على تطوير عدة أنظمة حاسوبية تطبيقية بكلية الملك فهد الأمنية.



- القيام بدراســة تقييم تطبيقات أنظمة المعلومات من حيث النجاح والفشــل ومن بينها نظام الإسعاف بمدينة لندن وكارثة (LASDS) عام ١٩٩٢م.
- دراسة تحليلية لتطبيق نظام ERP لإدارة عمل جمعية الهلال الأحمر التركية (TRCS).
- القيام بإعداد دورة تدريبية خاصة لطلبة الدكتوراة في قسم IRIS بجامعة سالفورد فيما يخص العمل على نظام التحليل الوصفي للبيانات NVivo.
- القيام بإعداد دورة تدريبية خاصة لطلبة الدكتوراة في قسم IRIS بجامعة سالفورد فيما يخص العمل على نظام الفهرسة الإلكترونية EndNote.





تم التصميم والإخراج الفني والطباعة في
الإدارة العامة للطباعة والنشر بمعهد الإدارة العامة - ١٤٣٩هـ

هذا الكتاب

يهدف كتاب ثورة البيانات إلى تقديم [illegible] مبدئي ودقيق للبيانات وثورة البيانات الجارية حالياً. فالكتاب يسعى، من جهة، لبيان بعض الطرق والوسائل المتعلقة بإعادة تشكيل إنتاج ومعالجة وتحليل وتبادل البيانات، وما يعني هذا لكيفية إنتاج واستخدام المعلومات والمعرفة؛ وهو، من جهة أخرى، يُمهد للنقاش والتفكير النقدي حول البيانات من حيث طبيعتها، وكيفية تأطيرها من النواحي الفنية والفلسفية والأخلاقية والاقتصادية والتجميعات التقنية والمؤسسية المحيطة بها. ويستند التحليل المقدم في هذا الكتاب إلى مشاركة واسعة النطاق وإلى ما سبقه من مؤلفات في مختلف العلوم، والعلوم الإنسانية، والعلوم الاجتماعية، ومن مطبوعات الثقافة الشعبية، والإصدارات الصحافية، والدوريات الصناعية؛ وإلى خبرة مباشرة في العمل على أنظمة حفظ أرشيفية وبنى تحتية ومشاريع تحليلات لبيانات على نطاق مؤسسي واسع. وينقسم الكتاب إلى أحد عشر فصلاً، يقدم الفصل الأول لمحة عامة ونقداً فكرياً حول مفهوم البيانات وكيفية فهم قواعد البيانات والبنى التحتية للبيانات، في حين يبحث الفصل الثاني في الدور المستمر للبيانات الصغيرة والكيفية التي تمّ بها توسيع نطاقها إلى أرشيفات/محفوظات رقمية وبنى تحتية، إلى أن وصلت إلى درجة التداول وبيعها من خلال وسطاء البيانات. كما يناقش الفصل الثالث الدافع للتوجه نحو استحداث البيانات المفتوحة والبيانات المرتبطة التي يجري تقاسمها ومشاركتها وإعادة استخدامها على نطاق أوسع بمرور الوقت. ويبيّن الفصلان الرابع والخامس طبيعة البيانات الكبيرة بالتفصيل وعوامل التمكين للبيانات الكبيرة ومصادرها المساعدة على انتشارها. ويقدّم الفصل السادس لمحة عامة عن مجموعة جديدة من تحليلات البيانات التي ترمي إلى فهم البيانات الصغيرة الموسعة والبيانات الكبيرة. كما يدرس الفصلان السابع والثامن الحجج المستخدمة لتعزيز البيانات الكبيرة والترويج لها وبيان تأثيرها في الحوكمة والأعمال، والسبل التي تقوم بها ثورة البيانات لإعادة تشكيل كيفية تصوّر البحوث وممارستها. ويناقش الفصلان التاسع والعاشر التحديات التقنية والتنظيمية، والأخلاقية، والسياسية، والقانونية لثورة البيانات. ويحدد الفصل الأخير بعض الاستنتاجات الكلية ويقدّم خارطة طريق لإجراء المزيد من البحث والتأصيل.



تصميم وإخراج وطباعة الإدارة العامة للطباعة والنشر
معهد الإدارة العامة ١٤٣٩هـ